used, but states that there remains no description of specific embodiments of the claimed invention other than ones featuring *E. coli*. The Office Action states that "no description is provided, for example, for tRNA genes corresponding to rarely used codons in plant cells or protozoa, two large classes of cell types embraced by the claims (i.e., rarely used codons)." The Office Action also states that the prior art does not appear to provide teachings as to which of the many known tRNA genes correspond to "rarely used" codons for the many different cell types encompassed by the claims, and that there is *no evidence of record that rare codon patterns have been established for a sufficient number of cell types for one of skill in the art to be able to envision a sufficient number of specific embodiments of the invention to describe the very broadly claimed genus*. Finally, with respect to Written Description, the Office Action states that there remains *no evidence of record to indicate that a sufficient number of tRNA genes obtained from different cell types corresponding to rarely used codons of different cell types were known in the prior art* for one of skill in the art to envision a sufficient number of embodiments of the claimed vectors and host cells to describe the broad genus of host cells and vectors encompassed by the claims. Applicants respectfully disagree.

Under the law, in order to satisfy the written description requirement, a patent specification must describe the claimed invention in sufficient detail that one skilled in the art can reasonably conclude that the inventor had possession of the claimed invention. *Vas-Cath, Inc. v. Mahurkar*, 935 F.2d 1555, 1563; M.P.E.P. §2163. Further, under *both* the written description and enablement requirements, one need not describe in detail that which is well known in the art. M.P.E.P §2163, citing *Hybritech v. Monoclonal Antibodies*, 802 F.2d 1367, 1384. The Office Action states that "the critical elements of applicant's invention are the tRNA genes corresponding to rarely used codons on the claimed vectors which are determined by the combination of host cell type (i.e., rarely used codons), the corresponding tRNA genes and the protein to be expressed." Applicants submit that patterns of codon usage were well known to those skilled in the art at the time of filing. For example, Nakamura et al., 1996, Nucleic Acids Res. 24: 214-215 (Exhibit A) provides codon usage tabulated from the GenBank international DNA sequence databases for 4,805 species. Applicants submit that these species include prokaryotes, protozoa and fungi, and a wide variety of higher eukaryotes, including animals and plants. The codon usage tabulation of Nakamura necessarily details low, as well as high codon usage in each of the thousands of species examined.

Further, Zhang et al., 1991, Gene 105: 61-72 (Exhibit B) details low usage codons in species as diverse as E. coli (prokaryotic), yeast (S. cerevisiae; protozoan), Drosophila and eleven species of primates. The paper describes the unique combination of least-used codons for the species examined.

In addition, Saier (1995, FEBS Lett. 362: 1-4; Exhibit C) describes rare codon usage in species including *Rhodobacter capsulatus*, *R. spheroides*, *Clostridium acetobutylicum*, *Streptomyces coelicolor* and *E. coli*. Saier relates the rare codon usage to the regulation of metabolically sensitive genes.

In view of these sources of information, particularly in view of the comprehensive nature of the Nakamura database before the filing of the subject application, Applicants submit that a sufficient number of rare codon usage patterns have been established in the prior art for a sufficient number of cell types for one of skill in the art to readily envision a sufficient number of specific embodiments of rare codon usage patterns to describe the claimed genus.

With regard to tRNA genes, Applicants submit that a wide variety of tRNA genes was also known in the art at the time of filing. For example, as early as 1984, Sprinzl & Gauss described a compilation of 353 sequences of tRNA genes including cellular and mitochondrial tRNAs from bacteria and phage, plants, yeasts and fungi, insects, amphibians and mammals, including rats, mice, cows and humans (Nucleic Acids Res. 12 Suppl.: r59-131; Exhibit D). Further, there was available on the World Wide Web as of the end of 1998 (before filing), a compilation of tRNA sequences and genes including 3279 sequences. This number is taken from documentation on the current WWW compilation at www.uni-bayreuth.de/departments/biochemie/trna (see Exhibit E). The thousands of tRNA genes described include those for rarely used tRNAs. Additional rare tRNAs are described in, for example: Kawakami et al., 1993, Genetics 135: 309-320 (Exhibit F), which describes a rare Arg-tRNA-CCU in *Saccharomyces cerevisiae*; and Clouthier et al., 1998, J. Bacteriol. 180: 840-845 (Exhibit G), which describes the rare Arg-tRNA-AGA in *Salmonella enteritidis*. Applicants therefore submit that there was known in the art, at the time of filing, a large number of tRNA genes, including those encoding rare tRNAs, from a broad cross section of species.

The Office Action states that the prior art does not appear to provide teachings as to which of the many known tRNA genes correspond to "rarely used" codons for the many different cell types encompassed by the claimed invention. Applicants submit that given the extensive data on codon usage available in the art (e.g., Exhibits A-C), one of skill in the art would know if a given tRNA gene, e.g., one described in any of Exhibits D, F or G, corresponds to a rarely used codon.

In view of the above, and given the description provided in the specification, Applicants submit that the invention of claims 1-16 and 18-44 is described in sufficient detail to enable one of skill in the art to envision a sufficient number of embodiments of the claimed vectors and host cells to describe the full scope of the claimed genus of vectors and host cells. Applicants respectfully request the withdrawal of the §112, first paragraph written description rejection of claims 1-16 and 18-44.

Rejections under 35 U.S.C. §103:

All claims remain rejected under 35 U.S.C. §103 as obvious over Del Tito et al. in combination with one or more of Makoff et al., the 1997 Novagen catalog, and Wnendt. The Office Action rejects the evidence of commercial success as an objective indicator of non-obviousness because "there remains no meaningful background against which the sales figures presented in Paper No. 10 can be weighed to determine if the demonstrated sales are so indicative of commercial success as to make the claimed invention unobvious." The Office Action further states that there needs to be a showing that the commercial success is commensurate with the claimed invention, and that "A demonstration of commercial success for a couple of specific embodiments useful in E. coli cannot be considered as evidence of nonobviousness commensurate with the full, broadly claimed genus of host cells and vector of the instant invention." Applicants respectfully disagree.

First, Applicants submit that the law does not absolutely require evidence of market share in order for commercial sales of a product of an invention to be persuasive of nonobviousness. The Federal Circuit has held that in order to demonstrate non-obviousness, the commercial success of the product must be due to the merits of the claimed invention beyond what was readily available in the prior art. *Richdel, Inc., v. Sunspool Corp.*, 714 F.2d 1573 (Fed. Cir.

1983). With regard to whether the commercial success is due to the merits of the claimed invention, in *J.T. Eaton & Co., Inc. v. Atlantic Paste & Glue Co.* the Federal Circuit held that a primary showing of commercial success *limited to sales*, coupled with a demonstration that the commercial success of the product derives from the claimed invention and is attributable to something disclosed in the patent that was not readily available in the prior art is entitled to the presumption that that the commercial success of the product is attributable to the patented invention. *J.T. Eaton & Co., Inc. v. Atlantic Paste & Glue Co.*,106 F.3d 1563, 1571 (Fed. Cir.1997).

Applicants submit, and the Buchanan Declarations support the conclusion, that the products sold are embodiments of the claimed invention, falling plainly within the scope of claim 1. This claim requires a host cell containing a recombinant DNA molecule which comprises an array of three of more tRNA genes, wherein the tRNA genes correspond to codons that are rarely used in the host cell. As stated in Ms. Buchanan's first Rule 132 Declaration, *each* of the competent cell products for which sales figures are reported contains three tRNA genes corresponding to rarely used codons. Applicants submit that the prior art *does not* teach cells with three or more tRNA genes corresponding to codons that are rarely used in the host cell. Further, the first Buchanan Declaration states that the commercial success of the claimed invention is not the result of heavy promotion or advertising, noting that Stratagene spent no more on promotion or advertising of this product than it did on any other competent cell product it sells. The product also sells for a considerably higher price than non-codon-enhanced cells sold by the same company. Thus, the commercial success of the product is attributable not to heavy promotion or lower price, but to something disclosed in the patent that was not readily available in the prior art.

Under *J.T. Eaton*, having shown the necessary correspondence between the commercial success and the claimed invention, Applicants are thus entitled to the presumption that the commercial success of the product is attributable to the claimed invention. Under these circumstances, also in accord with *J.T. Eaton*, Applicants submit that sales alone, in the absence of market data, are indicative of non-obviousness. In view of this, and, where, as in the instant case, there are no data regarding market share because there were no competing products at the

time of the sales reported (itself a strong indicator of non-obviousness), Applicants submit that the sales figures provided are evidence of non-obviousness.

With regard to the assertion that "A demonstration of commercial success for a couple of specific embodiments useful in E. coli cannot be considered as evidence of nonobviousness commensurate with the full, broadly claimed genus," Applicants submit that the prior art cited is drawn to expression in *E. coli*, as is the evidence of commercial success of the claimed invention. Thus, the scope of the demonstrated commercial success of the claimed invention is *directly relevant* to the non-obviousness of the claimed invention over the cited prior art. Thus, giving proper weight to the commercial success of the embodiments sold (as discussed above), *E. coli* embodiments of the claimed invention are non-obvious. Further, if *E. coli* embodiments, for which the prior art appears to be most relevant, are not obvious, Applicants submit that there is no reason to conclude that embodiments encompassing other cell types, for which there is a lack of relevant prior art, would also be non-obvious. Applicants therefore submit that the scope of the invention encompassed by the commercial embodiments is sufficient to overcome the alleged obviousness of the claimed invention.

In view of the above, Applicants submit that the invention of claims 1-16 and 18-44 is not obvious over the combination of references cited. Applicants respectfully request that the rejection of these claims under §103 be withdrawn.

Applicants submit that in view of the preceding remarks and the Exhibits provided, all issues raised in the Office Action have been addressed herein. Applicants respectfully request reconsideration of the claims.

Respectfully submitted,

Mark J. Fitzgerald
Reg. No. 45,928 for
Kathleen M. Williams

Dated: September 26, 2002

Kathleen M. Williams, Ph.D.
Registration No. 34,380
Attorney for Applicant
PALMER & DODGE LLP
111 Huntington Avenue
Telephone: (617) 239-0451
Telecopier: (617) 227-4420

EXHIBIT A

214–215 Nucleic Acids Research, 1996, Vol. 24, No. 1

Codon usage tabulated from the international DNA sequence databases

Yasukazu Nakamura, Ken-nosuke Wada[1], Yoshiko Wada, Hirofumi Doi[2], Shigehiko Kanaya[3], Takashi Gojobori and Toshimichi Ikemura*

National Institute of Genetics and The Graduate University for Advanced Studies, Mishima, Shizuoka 411, Japan, [1]ATR Limited, Seika, Kyoto 619-02, Japan, [2]Fujitsu Laboratories Ltd, Chiba 261, Japan and [3]Yamagata University, Yonezawa, Yamagata 992, Japan

Received August 31, 1995; Revised and Accepted October 4, 1995

ABSTRACT

Codon usage in 87 602 genes has been calculated using the nucleotide sequence data obtained from the GenBank Genetic Sequence Data Bank (Release 90.0; September 1995). The database is called the CUTG Database; the complete form of the database can be obtained by anonymous ftp from DDBJ and a part of the database, which lists the frequency of codon use in each organism, is made searchable through our World Wide Web server.

SOURCE AND METHODS

Codon usage in individual genes has been calculated using the nucleotide sequence data obtained from the GenBank Genetic Sequence Data Bank (Release 90.0; September 1995). The compilation of codon usage is synchronized with each major release of GenBank. The resulting database is called the CUTG database (1–5).

In selecting protein coding sequences we relied on the FEATURES tables of GenBank, and only complete genes without unambiguous bases were used in the analysis. In GenBank, a group of consecutive genes whose entire region had been sequenced were registered under one LOCUS name. To distinguish the different genes belonging to a single LOCUS, the symbol # followed by a number is added after the LOCUS name; the numbers represent the order of the CDS registered in the FEATURES table of GenBank. When introns of a gene have not been completely sequenced, some of its exons are registered in separate entries (LOCUS) in GenBank. These exons, belonging to the same gene but having different LOCUS names, were combined into one entry and the first LOCUS name is added.

For the biological significance of codon usage, see Ikemura (6) and Aota and Ikemura (7,8).

FILES

Files of the present database, containing codon usage of 87 602 CDSs of 4805 species, are available by anonymous ftp from DDBJ. Files named as gb***.codon list the codon use in each gene registered in the GenBank Sequence files (gb***.seq). The LOCUS names given in GenBank were used to designate individual genes. Each LOCUS name is followed by fields of information extracted from the FEATURES of each CDS for defining each open reading frame analyzed here. The order of the codons in the table is the same as the previous compilation (see the CODON_LABEL file or REFERENCES).

To reveal the characteristics of codon use of a wide range of organisms, as well as viruses and organella, the frequency (per 1000) of codon use in 461 organisms for which >20 genes are available was calculated by summing up numbers of codon used. World Wide Web clients, such as NCSA Mosaic and Netscape, may be used to query this file. A user can display a codon usage table by clicking an anchor for selecting species or searching with species' name (Fig. 1).

DISTRIBUTION AND ACCESS

Complete form of the database is available by anonymous ftp from DDBJ:

ftp://ftp.nig.ac.jp/pub/db/codon/GB90.

The file README contains the latest information on the database in plain text format.

The frequencies of codon use in 461 organisms for which >20 genes are available can be accessed on the following WWW server:

http://tisun4a.lab.nig.ac.jp/codon/CUTG.html.

Comments on the database can be sent to cutg@lab.nig.ac.jp by e-mail.

ACKNOWLEDGEMENTS

The authors wish to thank Dr Y. Ugawa at the DNA Information and Stock Center, National Institute of Agrobiological Resources, for help in constructing the database. This work was supported by a grant-in-aid for creative basic research (Human Genome Project) and for scientific research on a priority area (Genome Informatics) and by a grant-in-aid of scientific research from the Ministry of Education, Science and Culture of Japan.

* To whom correspondence should be addressed

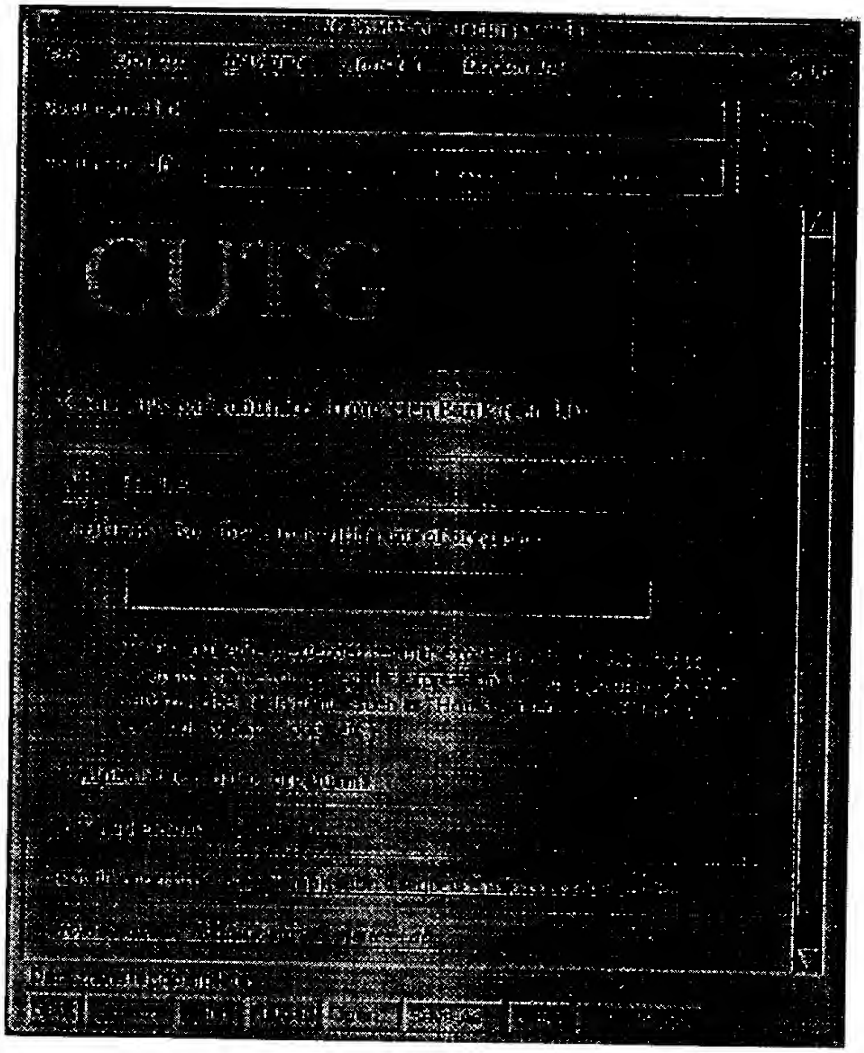


Figure 1. Snapshot of the CUTG home page.

REFERENCES

1 Maruyama, T., Gojobori, T., Aota, S. and Ikemura, T. (1986) *Nucleic Acids Res.*, **14**, r151–r197.
2 Aota, S. Gojobori, T., Ishibashi, F., Maruyama, T. and Ikemura, T. (1988) *Nucleic Acids Res.*, **16**, r315–r402.
3 Wada, K., Aota, S., Tsuchiya, R., Ishibashi, F., Gojobori. and Ikemura, T. (1990) *Nucleic Acids Res.*, **18**, 2367–2441.
4 Wada, K., Wada, Y., Doi, H., Ishibashi, F., Gojobori, T. and Ikemura, T. (1991) *Nucleic Acids Res.*, **19**, 1991–1986.
5 Wada, K., Wada, Y., Ishibashi, F., Gojobori, T. and Ikemura, T. (1992) *Nucleic Acids Res.*, **20**, 2111–2118.
6 Ikemura, T. (1985) *Mol. Biol. Evol.*, **2**, 13–24.
7 Aota, S. and Ikemura, T. (1986) *Nucleic Acids Res.*, **14**, 6345–6355.
8 Aota, S. and Ikemura, T. (1986) *Nucleic Acids Res.*, **14**, 8702.

Gene, 105 (1991) 61–72

EXHIBIT B

GENE 06001

Low-usage codons in *Escherichia coli*, yeast, fruit fly and primates

(Recombinant DNA; GenBank; codon bias; *Saccharomyces cerevisiae*; *Drosophila melanogaster*; *Homo sapiens*)

Shiping Zhang[a*], Geoffrey Zubay[a] and Emanuel Goldman[b]

[a] *Fairchild Center for Biological Sciences, Columbia University, New York, NY 10027 (U.S.A.) Tel. (212)854-4578; and* [b] *Department of Microbiology and Molecular Genetics, UMDNJ – New Jersey Medical School, Newark, NJ 07103 (U.S.A.)*

Received by W.M. Holmes: 12 September 1990
Revised/Accepted: 17 April/31 May 1991
Received at publishers: 18 June 1991

SUMMARY

Codon usage is compared between four classes of species, with an emphasis on characterization of low-usage codons. The classes of species analyzed include the bacterium *Escherichia coli* (ECO), the yeast *Saccharomyces cerevisiae* (YSC), the fruit fly *Drosophila melanogaster* (DRO), and several species of primates (PRI) (taken as a group; includes eleven species for which nucleotide sequence data have been reported to GenBank, however, greater than 90% of the sequences were from *Homo sapiens*). The number of protein-coding sequences analyzed were 968 for ECO, 484 for YSC, 244 for DRO, and 1518 for PRI. Three methods have been used to determine low-usage codons in these species. The first and most common way of assessing codon usage is by summing the number of time codons appear in reading frames of the genome in question. The second way is to examine the distribution of usage in different genes by scoring the number of protein reading frames in which a particular codon does *not* appear. The third way starts with a similar notion, but instead considers *combinations* of codons that are missing from the maximum number of genes. These three methods give very similar results. Each species has a unique combination of eight least-used codons, but all species contain the arginine codons, CGA and CGG. The agreement between YSC and PRI is particularly striking as they share six low-usage codons. All six carry the dinucleotide sequence, CG. The eight least-used codons in PRI include all codons that contain the CG dinucleotide sequence. Low-usage codons are clearly avoided in genes encoding abundant proteins for ECO, YSC and DRO. In all species, proteins containing a high percentage of low-usage codons could be characterized as cases where an excess of the protein could be detrimental. Low codon usage is relatively insensitive to gross base composition. However, dinucleotide usage can sometimes influence codon usage. This is particularly notable in the case of CG dinucleotides in PRI.

INTRODUCTION

Amino acids (aa) that are represented by more than one codon usually do not use synonym codons equally (e.g., Grantham et al., 1981; Gouy and Gautier, 1982). Indeed, the differential use of codons is most striking and species-specific and raises many interesting possibilities and concerns (Andersson and Kurland, 1990). Studies on ECO and YSC have shown that high abundance proteins show a sharp avoidance of codons that are in low usage in the overall gene population (Post et al., 1979; Ikemura, 1981; 1982; Bennetzen and Hall, 1982), a finding that has led to

Correspondence to: Dr. E. Goldman, Dept. of Microbiology, New Jersey Med. School, 185 S. Orange Ave., Newark, NJ 07103 (U.S.A.) Tel. (201)456-4367; Fax (201)456-3644.

* Current address: Biology Department, Brookhaven National Laboratory, Upton, NY 11973 (U.S.A.)

Abbreviations: aa (a.a. in Tables), amino acid(s); bp, base pair(s); DRO, *Drosophila melanogaster*; ECO, *Escherichia coli*; PRI, primates; r, ribosomal; STP, stop codon; YSC, yeast *Saccharomyces cerevisiae*.

the suggestion that low-usage codons may be more difficult to translate (Grosjean and Fiers, 1982; Konigsberg and Godson, 1983; Kurland, 1987). This suggestion is supported by the observation that cognate tRNA abundance is roughly proportional to codon usage for most codons (Ikemura, 1985) and that rates of translation can vary with concentrations of charged tRNA, at least in ECO (Rojiani et al., 1990). Direct assays of translatability lend some support to this view (Robinson et al., 1984; Bonekamp et al., 1985; Carter et al., 1986; Hoekema et al., 1987; Sørensen et al., 1989; Curran and Yarus, 1989; Chen and Inouye, 1990). Further, expression of at least one heterologous protein in ECO required oligodeoxynucleotide synthesis of the entire gene with high usage synonym codons of ECO, since message containing the naturally occurring codons of the gene (very rich in ECO's low-usage codons) failed to support measurable translation (Abate et al., 1990; T. Curran, personal communication).

Whereas papers summarizing codon usage have appeared sporadically (e.g., Grantham et al., 1981; Ikemura, 1985; Sharp and Li, 1986; Sharp et al., 1988; Wada et al., 1990), the amount of information keeps increasing at the rate of about 2×10^6 bp per year. This necessitates periodic reevaluation and updating of codon usage. There is also the need for new ways of looking at codon usage which requires new perspectives and new computer programs. For example, Gutman and Hatfield (1989) have considered the problem from the perspective of frequency of appearance of codon pairs.

For the purpose of this study and to stay within reasonable practical limits, we will confine the information presented in this paper to ECO, YSC, DRO, and PRI. Mitochondrial genomes are not included in this analysis. Here, we shall describe different ways of estimating codon usage for these four classes of organisms. In addition to serving as an update of previous papers, a new way of evaluating codon usage is introduced and the relationship between codon usage and dinucleotide frequency is addressed. A major purpose of this paper is to give an accurate assessment (as of September 1990) of low-usage codons in different species. A second and more elusive goal is to seek explanations for the evolutionary choices that have been made.

RESULTS AND DISCUSSION

(a) Criteria for identifying low-usage codons

We shall define codon usage as the number of times a codon is translated per unit time. This has not been measured directly in vivo, so estimates of codon usage are obtained by indirect observations. It should be appreciated that codon usage is likely to be different under different conditions of growth. For example, under conditions of rapid growth when there are plenty of nutrients, the overall rate of protein synthesis will be maximal and the synthesis of proteins that are needed for rapid growth should be favored. By contrast, under conditions of nutrient limitation, a new set of proteins is most likely to dominate metabolism and the overall rate of protein synthesis would be reduced (Ingraham et al., 1983). We will describe three ways of estimating codon usage. Each of them has its advantages and drawbacks. Fortunately, they all give approximately the same answer as regards the hierarchy for 'most used' and 'least used' codons within each synonymous codon family.

(1) Sums of codon appearance

The most common way of measuring codon usage is by summing the number of times codons appear in the reading frames of the genome. This approach is summarized in Table I for four different types of organisms under discussion. This method should over-estimate codons used infrequently and under-estimate codons used frequently. The reason for this is that, when averaging over the entire genome, no weight is given to the number of times different reading frames are used, which is reflected in the variable quantities of protein products. This weighting is a complex resultant of transcriptional efficiency, message stability, and translational efficiency. In general, this weighting is not precisely known but frequently it can be roughly estimated from the amount of gene-encoded protein product.

(2) Absence of particular codons in genes

A second way of estimating codon usage is to examine the distribution of usage in different genes. We have done this by scoring the number of protein reading frames in which a particular codon does *not* appear. In Table II, these data are directly compared for the four species under consideration. A large number indicates a narrow distribution for the codon in question. No attempt has been made to weight this method by scoring the number of times a codon appears in a gene. Only the presence or absence of a codon within a gene has been scored. This approach to measuring codon usage is based on the prediction that the most used codons should have the broadest distribution and the least used codons should have the narrowest distribution. The data in Table II are presented in a way that is most convenient for making comparisons of relative usage between different organisms. For this purpose, the numbers have been normalized to those actually observed for DRO. The normalization factors are given in the footnote of Table II.

(3) Combinations of excluded codons

The third way in which we have estimated codon usage starts with a similar notion to that used in the second

TABLE I

A comparison of codon usages[a] among four different species[b]

A. ECO: total codons: 323059, total proteins: 968

	U	C	A	G	
U	8089 (Phe)	3382 (Ser)	4875 (Tyr)	1552 (Cys)	U
	5838 (Phe)	3048 (Ser)	4295 (Tyr)	1961 (Cys)	C
	3397 (Leu)	2105 (Ser)	644 (STP)	859 (STP)	A
	3659 (Leu)	2549 (Ser)	65 (STP)	4169 (Trp)	G
C	3204 (Leu)	2123 (Pro)	3666 (His)	7031 (Arg)	U
	3135 (Leu)	1354 (Pro)	3471 (His)	6946 (Arg)	C
	960 (Leu)	2622 (Pro)	4223 (Gln)	989 (Arg)	A
	17477 (Leu)	7734 (Pro)	9589 (Gln)	1491 (Arg)	G
A	8609 (Ile)	5499 (Thr)	5265 (Asn)	2380 (Ser)	U
	8712 (Ile)	7872 (Thr)	7868 (Asn)	4331 (Ser)	C
	1272 (Ile)	2108 (Thr)	12104 (Lys)	592 (Arg)	A
	8507 (Met)	4050 (Thr)	3857 (Lys)	428 (Arg)	G
G	6716 (Val)	5770 (Ala)	10382 (Asp)	9202 (Gly)	U
	4652 (Val)	7490 (Ala)	7116 (Asp)	9824 (Gly)	C
	3896 (Val)	6760 (Ala)	14134 (Glu)	2246 (Gly)	A
	7972 (Val)	10641 (Ala)	6149 (Glu)	3110 (Gly)	G

B. YSC: total codons: 241099, total proteins: 484

	U	C	A	G	
U	5559 (Phe)	5954 (Ser)	3962 (Tyr)	1842 (Cys)	U
	4813 (Phe)	3535 (Ser)	3898 (Tyr)	900 (Cys)	C
	3887 (Leu)	3722 (Ser)	255 (STP)	136 (STP)	A
	7776 (Leu)	1565 (Ser)	92 (STP)	2490 (Trp)	G
C	2297 (Leu)	3083 (Pro)	2977 (His)	1801 (Arg)	U
	981 (Leu)	1388 (Pro)	2017 (His)	467 (Arg)	C
	2840 (Leu)	5162 (Pro)	7098 (Gln)	531 (Arg)	A
	2003 (Leu)	988 (Pro)	2500 (Gln)	255 (Arg)	G
A	7454 (Ile)	5281 (Thr)	7587 (Asn)	2802 (Ser)	U
	4494 (Ile)	3425 (Thr)	6235 (Asn)	1782 (Ser)	C
	3054 (Ile)	3703 (Thr)	9104 (Lys)	5791 (Arg)	A
	5130 (Met)	1587 (Thr)	8536 (Lys)	1814 (Arg)	G
G	6486 (Val)	6884 (Ala)	8962 (Asp)	8376 (Gly)	U
	3622 (Val)	3765 (Ala)	5418 (Asp)	2146 (Gly)	C
	2366 (Val)	3666 (Ala)	11336 (Glu)	2137 (Gly)	A
	2280 (Val)	1242 (Ala)	4082 (Glu)	1228 (Gly)	G

C. DRO: total codons: 125527, total proteins: 244

	U	C	A	G	
U	1473 (Phe)	776 (Ser)	1337 (Tyr)	710 (Cys)	U
	2976 (Phe)	2541 (Ser)	2514 (Tyr)	1782 (Cys)	C
	589 (Leu)	772 (Ser)	127 (STP)	50 (STP)	A
	1831 (Leu)	2121 (Ser)	67 (STP)	1258 (Trp)	G
C	924 (Leu)	719 (Pro)	1304 (His)	1267 (Arg)	U
	1500 (Leu)	2393 (Pro)	2135 (His)	2217 (Arg)	C
	313 (Leu)	1542 (Pro)	1680 (Gln)	873 (Arg)	A
	4606 (Leu)	1982 (Pro)	4670 (Gln)	908 (Arg)	G
A	2001 (Ile)	1088 (Thr)	2608 (Asn)	1215 (Ser)	U
	3237 (Ile)	2939 (Thr)	3318 (Asn)	2369 (Ser)	C
	882 (Ile)	1207 (Thr)	1902 (Lys)	629 (Arg)	A
	2997 (Met)	1704 (Thr)	5923 (Lys)	777 (Arg)	G
G	1371 (Val)	1900 (Ala)	3385 (Asp)	1983 (Gly)	U
	1958 (Val)	4797 (Ala)	3199 (Asp)	3824 (Gly)	C
	885 (Val)	1435 (Ala)	2308 (Glu)	2451 (Gly)	A
	3424 (Val)	1714 (Ala)	5773 (Glu)	599 (Gly)	G

D. PRI: total codons: 606694, total proteins: 1518

	U	C	A	G	
U	9684 (Phe)	8052 (Ser)	7504 (Tyr)	5982 (Cys)	U
	13718 (Phe)	10577 (Ser)	11195 (Tyr)	8522 (Cys)	C
	3279 (Leu)	5756 (Ser)	442 (STP)	774 (STP)	A
	8670 (Leu)	2448 (Ser)	302 (STP)	8290 (Trp)	G
C	6538 (Leu)	9352 (Pro)	5843 (His)	2818 (Arg)	U
	12177 (Leu)	12138 (Pro)	8815 (His)	6945 (Arg)	C
	3871 (Leu)	9007 (Pro)	8661 (Gln)	3258 (Arg)	A
	26255 (Leu)	3901 (Pro)	20956 (Gln)	8354 (Arg)	G
A	9159 (Ile)	7770 (Thr)	10115 (Asn)	5747 (Ser)	U
	14619 (Ile)	13820 (Thr)	18569 (Asn)	11285 (Ser)	C
	3605 (Ile)	8768 (Thr)	13636 (Lys)	6086 (Arg)	A
	13572 (Met)	4064 (Thr)	21516 (Lys)	6627 (Arg)	G
G	8238 (Val)	11788 (Ala)	13231 (Asp)	6627 (Gly)	U
	9773 (Val)	17754 (Ala)	17615 (Asp)	15198 (Gly)	C
	3810 (Val)	8514 (Ala)	18379 (Glu)	10442 (Gly)	A
	19009 (Val)	4509 (Ala)	25461 (Glu)	10497 (Gly)	G

[a] DNA sequences analyzed were from GenBank version 63 (15 March 1990). Reading frames were based on 'pept' labels (or 'CDS' labels in the current GenBank format) in the FEATURE sections. Repeating sequences were excluded. Computer programs used are written in C and run on a Sun3.

[b] The species analyzed include the bacterium *E. coli* (ECO), the yeast *S. cerevisiae* (YSC), *D. melanogaster* (DRO), and several species of primates (PRI) (taken as a group; includes eleven species for which DNA sequence data have been reported, however, greater than 90% of the sequences were from man).

method but instead considers *combinations* of codons that are missing from the maximum number of genes. The assumption here is that combinations of low-usage codons should be excluded from the most active genes, because the more low-usage codons present in a gene, the greater the potential reduction in the amount of translated product. Although this remains to be experimentally tested in a rigorous fashion, there are a number of reasons to predict this result (see INTRODUCTION). The program searches for combinations of codons (two or more) that are excluded from the maximum number of reading frames of the species in question. To make the computer calculations practical, only the 20 least used codons defined by method 2 above were used in the pool to determine combinations of least-used codons. In Table III, results are presented for combinations of up to eight codons with the least favored incidence; the number of protein reading frames that exclude the combination is given at the left in the table. Frequently, there is little difference between first, second and third (not shown) choices, indicating a degree of uncertainty in giving priority to the choices.

For ECO, there is a stepwise addition of a new low-usage codon at each stage as the size of the combination increases. This is also true for YSC except after stage 7 (combinations containing seven codons) where GGG is eliminated and AGG and ACG are simultaneously added. For DRO, a similar stepwise pattern of addition is seen as for ECO: The situation is far more complex and intriguing for PRI. The

TABLE II

Relative numbers[a] of proteins not using a particular codon in four species

codon (a.a.)	ECO	YSC	DRO	PRI	codon (a.a.)	ECO	YSC	DRO	PRI
UUU (Phe)	17	15	39	35	AUU (Ile)	10	4	18	38
UUC (Phe)	17	6	12	14	AUC (Ile)	10	8	17	14
UUA (Leu)	46	8	125	110	AUA (Ile)	124	55	64	88
UUG (Leu)	37	4	28	38	AUG (Met)	1	0	0	0
UCU (Ser)	33	5	49	28	ACU (Thr)	28	6	50	28
UCC (Ser)	42	11	14	10	ACC (Thr)	16	14	13	9
UCA (Ser)	68	30	68	40	ACA (Thr)	66	39	61	27
UCG (Ser)	80	59	18	94	ACG (Thr)	36	63	46	50
UAU (Tyr)	30	22	37	23	AAU (Asn)	25	19	22	33
UAC (Tyr)	30	5	19	16	AAC (Asn)	12	4	11	9
UAA (STP)	82	116	117	173	AAA (Lys)	4	3	33	17
UAG (STP)	228	198	177	195	AAG (Lys)	25	2	9	2
UGU (Cys)	88	40	75	48	AGU (Ser)	65	40	69	40
UGC (Cys)	76	89	19	38	AGC (Ser)	20	51	18	14
UGA (STP)	179	175	194	120	AGA (Arg)	161	7	95	45
UGG (Trp)	43	31	39	28	AGG (Arg)	180	56	67	23
CUU (Leu)	41	48	48	35	GUU (Val)	10	6	23	38
CUC (Leu)	39	35	20	12	GUC (Val)	23	12	19	16
CUA (Leu)	127	25	69	59	GUA (Val)	19	52	69	79
CUG (Leu)	9	54	15	5	GUG (Val)	17	43	15	7
CCU (Pro)	56	25	54	21	GCU (Ala)	15	7	24	10
CCC (Pro)	98	65	17	16	GCC (Ala)	18	10	3	4
CCA (Pro)	49	7	27	34	GCA (Ala)	11	30	40	18
CCG (Pro)	20	87	31	73	GCG (Ala)	13	73	28	64
CAU (His)	38	25	41	46	GAU (Asp)	11	6	14	18
CAC (His)	41	28	25	21	GAC (Asp)	16	5	17	5
CAA (Gln)	28	5	52	43	GAA (Glu)	7	2	23	17
CAG (Gln)	10	48	11	4	GAG (Glu)	19	25	7	4
CGU (Arg)	19	37	31	81	GGU (Gly)	13	6	20	32
CGC (Arg)	21	128	20	47	GGC (Gly)	14	34	11	10
CGA (Arg)	124	144	67	68	GGA (Gly)	68	47	15	25
CGG (Arg)	110	167	72	45	GGG (Gly)	49	70	80	18

[a] Normalization factor for each species is the ratio of total DRO proteins to the total proteins of the species. Total proteins used are (normalization factors in parentheses): DRO, 244 (1); ECO, 968 (0.252); YSC, 484 (0.504); PRI, 1518 (0.161). To obtain the actual (instead of relative) number of proteins in each species not using a particular codon, multiply the numbers in the table by the reciprocal of the normalization factor. For example, out of the 968 ECO proteins analyzed, the actual number of ECO proteins not using the CUA codon is 127 (from the table) times 3.968 (reciprocal of normalization factor 0.252) = 504.

four codons at stage 4 are eliminated at stage 5 never to reappear as part of the favored low-usage combination at higher stages. It should be noted that these four codons are all of the UA type. At stage 5, all four codons ending in CG emerge along with one codon, CGU, carrying a CG in the first two positions. At all future stages, codons containing

TABLE III

Numbers of proteins not using a codon or combination of codons[a]

A. ECO total proteins: 968

714	AGG
536	AGG AGA
352	AGG AGA AUA
255	AGG AGA CUA AUA
203	AGG AGA CUA AUA CGA
159	AGG AGA CUA AUA CGA CGG
129	AGG AGA CUA AUA CGA CGG CCC
83	AGG AGA CUA AUA CGA CGG CCC UCG

B. YSC total proteins: 484

352	CGG
243	CGG CGA
178	CGG CGA CGC
136	CGG CGA CGC CCG
115	CGG CGA CGC CCG GUC
95	CGG CGA CGC CCG CUC GCG
80	CGG CGA CGC CCG CUC GCG GGG
72	CGG CGA CGC CCG CUC GCC ACG AGG

C. DRO total proteins: 244

125	UUA
46	UUA AGA
70	UUA AGA AUA
55	UUA AGA GGG AUA
40	UUA AGA GGG AUA CGG
35	UUA AGA AUA CGC AGU CGA
30	UUA AGA GGG AUA CGG AGU CGA
25	UUA AGA GGG AUA UGU CGG AGU CGA

D. PRI total proteins: 1518

687	UUA
380	UUA AUA
219	UUA AUA GUA
116	UUA AUA GUA CUA
80	UCG CGU CCG GCG AGG
60	CGA CGU CCG CGC AGG CGG
51	UCG CGU CCG CGA AGG CGC CGG
43	UCG CGU CCG CGA AGG CGC CGG GCG

[a] The indicated codons are absent from the listed number of proteins on the same line, for each species.

a CG in either their first two or last two positions dominate. At stage 8, we find all of the codons that carry the dinucleotide sequence CG.

From the standpoint of computer programming this third method of picking low-usage codons is the most complex. However, with the program in hand it is easy to apply. For purposes of comparison, in most of the remainder of this paper we will use combinations of eight low-usage codons determined by the third method. These codons for the different species are presented together in Table IV. Each species has a unique combination of eight codons, but all species contain the Arg codons CGA and CGG. The agreement between YSC and PRI is particularly striking as they share six low-usage codons. All six carry the dinucleotide sequence CG. It is appropriate to mention at this point that CG is the dinucleotide sequence that is most avoided in PRI. This may be related to the fact that the C in a CG sequence is susceptible to methylation (e.g., see Razin and Riggs, 1980) and is therefore reserved for special situations. In YSC there is no methylation of C residues in CG sequences. Despite this we shall see below that this dinucleotide is also the most unpopular in YSC.

(b) Large codon families are more likely to contain low-usage codons as modulators of expression

There is reason to believe that, if codon usage plays a role in modulating gene expression, it is more likely to do so in the larger codon families that contain three of more synonymous codons. Two aa, Trp and Met, are represented by single codons. The two-codon families include Phe, Tyr, His, Gln, Asn, Lys, Asp, Glu and Cys. In most of these cases, both codons are believed to be capable of recognition by a single cognate tRNA (or anticodon) in ECO (Sprinzl et al., 1987, and references therein). This does not exclude the possibility that some of these codons are used as modulators of gene expression for example, because of the different strengths of anticodon–codon interactions (Grosjean et al., 1978; Yarus et al., 1986). In the cases of all the remaining codon families, where three, four or six codons are involved, at least two tRNAs are found for every aa. Most frequently in ECO and YSC (sufficient information is not available for DRO or PRI), as indicated above, there is a correlation between the abundance of the tRNA and the abundance of the codon (Ikemura, 1981; 1982). There are notable exceptions as in the case of the CGA codon for Arg. This codon is recognized in ECO by the same tRNA that recognizes the two high-usage codons CGU and CGC (Murao et al., 1972). It seems reasonable that we should focus our attention on the major codon families to gain an appreciation of the significance of low-usage codons.

In Table V, we have compared the three methods for assessment of codon usage for the nine major codon families. In this table, codons representing the same aa are grouped together. For each species, the % is given for codon usage determined by abundance in the genome as in Table I. In parentheses, the normalized number is given of proteins of the species that do not contain the codon as in Table II. A large number in parentheses should correspond to a small % if the two methods are consistent. Inspection of Table V indicates that the two methods are in quite good agreement for most aa. With only five exceptions (Val and Ala in ECO, Val in YSC, and Ser and Thr in DRO), the least abundant codon is excluded from the most proteins. With only five exceptions (Val, Ala and Gly in ECO, and Ser and Arg in PRI), the most abundant codon within the aa codon family is excluded from the lowest numbers of proteins; that is, it has the broadest distribution. In Table V, the % use of codons designated as low-usage codons by the combination of excluded-codons methods described above (see Tables III and IV) are underlined. Only in DRO do we have a codon designated by this latter method that does not appear in Table V (the UGU codon for Cys). Simple inspection of Table V suggests that additional codons might be designated as low-usage codons. However, we have chosen to focus our attention on eight codons from each species which we think are most likely to represent *bona fide* low-usage codons. If we were to attempt to expand this list we would run the risk of selecting codons that may not possess the most important properties of low-usage codons. Until such time as reliable quantitative estimates on rate of translatability can be given, we prefer to make comparisons

TABLE IV

Low-usage codons[a]

ECO	YSC	DRO	PRI	a.a.
AGG	AGC			Arg
AGA		AGA		Arg
AUA		AUA		Ile
CUA				Leu
CGA	CGA	CGA	CGA	Arg
CGG	CGG	CGG	CGG	Arg
CCG				Pro
UCG			UCG	Ser
	CGC		CGC	Arg
	CCG		CCG	Pro
	CUC			Leu
	GCG		GCG	Ala
	ACG		ACG	Thr
		UUA		Leu
		GGG		Gly
		AGU		Ser
		UGU		Cys
			CGU	Arg

[a] The eight least used codons for each species, as determined by the 'combinations of excluded codons' method (see text and Table III).

with a smaller subset of low-usage codons that we can be relatively sure of.

TABLE V

Synonymous codon usage for major codon families

		ECO		YSC		DRO		PRI	
Leu	UUA	11[a]	(46)[b]	27	(8)	6[c]	(125)	6	(110)
	UUG	11	(37)	36	(4)	18	(26)	11	(38)
	CUU	10	(41)	11	(48)	9	(48)	11	(35)
	CUC	10	(39)	5	(86)	15	(20)	21	(12)
	CUA	3	(127)	13	(25)	8	(69)	7	(59)
	CUG	55	(9)	9	(54)	44	(15)	45	(5)
Ile	AUU	47	(10)	50	(4)	33	(18)	33	(58)
	AUC	46	(10)	30	(8)	53	(17)	53	(14)
	AUA	7	(124)	20	(55)	14	(84)	13	(88)
Val	GUU	29	(10)	44	(6)	19	(22)	18	(38)
	GUC	20	(23)	25	(12)	26	(19)	25	(13)
	GUA	17	(19)	16	(52)	9	(68)	9	(79)
	GUG	34	(17)	15	(43)	46	(15)	49	(7)
Ser	UCU	18	(33)	31	(5)	8	(49)	18	(26)
	UCC	17	(42)	18	(11)	26	(14)	24	(10)
	UCA	12	(68)	19	(30)	8	(63)	13	(49)
	UCG	14	(60)	8	(69)	22	(18)	6	(94)
	AGU	13	(65)	15	(40)	12	(59)	13	(40)
	AGC	26	(20)	9	(51)	24	(18)	26	(14)
Pro	CCU	15	(56)	29	(26)	11	(54)	27	(21)
	CCC	10	(93)	13	(66)	35	(17)	35	(16)
	CCA	19	(49)	49	(7)	23	(27)	26	(34)
	CCG	56	(20)	9	(97)	30	(31)	11	(73)
Thr	ACU	20	(29)	38	(6)	16	(50)	23	(28)
	ACC	45	(16)	24	(14)	42	(13)	40	(9)
	ACA	12	(68)	26	(39)	17	(61)	25	(27)
	ACG	23	(36)	11	(63)	25	(46)	12	(80)
Ala	GCU	19	(15)	44	(7)	19	(24)	28	(10)
	GCC	24	(18)	24	(10)	49	(3)	42	(4)
	GCA	22	(11)	24	(30)	15	(40)	20	(18)
	GCG	35	(13)	8	(73)	17	(28)	11	(84)
Arg	CGU	43	(19)	17	(37)	19	(31)	9	(81)
	CGC	37	(21)	4	(128)	33	(20)	22	(47)
	CGA	5	(124)	5	(144)	13	(67)	10	(88)
	CGG	8	(110)	2	(157)	14	(72)	20	(45)
	AGA	4	(161)	54	(7)	9	(95)	19	(45)
	AGG	2	(180)	17	(58)	12	(67)	21	(23)
Gly	GGU	38	(13)	60	(8)	22	(20)	15	(32)
	GGC	40	(14)	15	(34)	43	(11)	36	(10)
	GGA	9	(68)	15	(47)	28	(15)	24	(25)
	GGG	13	(49)	9	(70)	7	(88)	25	(18)

[a] Numbers in each column indicate % use of synonymous codon (derived from Table I). For example, in ECO we counted a total of 31832 Leu codons of all types in Table I. Of these, 17477 were CUG codons; thus, the % use of CUG is 17477 ÷ 31832 × 100 = 55% (in sixth line).

[b] Numbers in parentheses indicate normalized number of proteins lacking the codon (taken directly from Table II).

[c] Underlined are numbers of low-usage codons determined by 'combinations of excluded codons' method (see Tables III and IV).

(c) Low-usage codons are preferentially avoided in genes for abundant proteins in ECO, YSC and DRO

Implicit in the consideration of low-usage codons is the notion that such codons are more slowly translated and therefore avoided in genes for proteins required in large amounts. This notion is supported by a comparison of proteins containing small and large numbers of low-usage codons (e.g., Ikemura, 1985). For combinations of eight excluded codons, the distribution of proteins according to the number of low-usage codons they contain is given in Table VI.

It is informative to inspect those proteins that fall into the 'zero class', which uses none of the eight designated low-usage codons, and perhaps equally informative to inspect those proteins at the other extreme that use an abnormally high concentration of these low-usage codons. In spite of the fact that the precise molar amounts of each of these proteins are not available, the trends are unmistakable. Thus for the 93 proteins that are in the zero class for ECO, 35 r-proteins are found. Most of the remainder are major proteins involved in protein synthesis initiation and elongation, enzymes involved in intermediary aa metabolism and carbohydrate metabolism and the major outer membrane proteins, OmpR and OmpF. In YSC the zero class also abounds with r-proteins. The zero class in YSC also contains some histones, enzymes involved in aa biosynthesis and fermentation, actin, alcohol dehydrogenase and several ubiquitins. Of the 25 proteins in the DRO zero class the same trend continues with a strong representation of major proteins or at least proteins that are major in some tissues some of the time. Thus, we see some r-proteins, metallothionein, myosin, major heat-shock proteins, alcohol dehydrogenase I, cytochrome *c*, tropomyosin and a tubulin appearing in the zero class. Only in PRI is the trend unclear. Thus, we do not find r-proteins, histones or major enzymes in the zero class. Rather we find some globins, interferons, interleukins and metallothionein. It may be that these proteins are needed in large amounts in special situa-

TABLE VI

Distributions of proteins with respect to the percentage of low-usage codons they contain

% of low codons	ECO	YSC	DRO	PRI
0%	93[a]	72	25	43
above 0% to 3%	443	168	48	307
above 3% to 6%	278	190	94	539
above 6% to 9%	102	46	48	862
above 9% to 12%	30	5	16	155
above 12% to 15%	12	2	8	73
above 15%	10	1	5	39

[a] Numbers refer to number of proteins containing the indicated % of low usage codons.

tions and therefore it is beneficial for their translation to avoid the designated low-usage codons. Alternately, it may be that translatability plays a smaller role in the selection of low-usage codons in PRI than it does in the other species considered here.

(d) Proteins that contain a high percentage of low-usage codons in their genes belong to classes where an excess of the protein could be detrimental

It has been argued that only high-usage codons are selected in abundant proteins, and there is no selection for low-usage codons in low-expression-level proteins (Sharp and Li, 1986; Andersson and Kurland, 1990). Nevertheless, an examination of the trends seen in classes with high frequency of low-usage codons raises the possibility of selection for low-usage codons. For example, reading frames found on transposable elements abound in classes with high frequency of low-usage codons. Indeed, transposition of the YSC Ty1 transposon has been reported to be regulated by the concentration of the cognate tRNA for one of YSC's low-usage codons, AGG (Xu and Boeke, 1990). It seems likely that transposition would be under very tight control in most species. The remainder of the proteins found in the ECO list with greater than 9% low-usage codons include some toxins and lesser known regulatory proteins. YSC in the 6% or greater low-usage class contains several regulatory protein genes and nuclear genes that encode products for mitochondria. The DRO proteins with greater than 9% low-usage codons are most notable for the large number of transposable element genes. They also contain the Shaker genes which are believed to be essential for potassium channels in the nervous system (Pongs et al., 1988). The PRI with greater than 12% low-usage codons abound in proto-oncogenes, growth factor genes, several hormone genes and, strangely enough, several histone genes. The finding of histone genes in the class with high frequency of low-usage codons in PRI contrasts with the findings in YSC where several histone genes are found in the zero class. It may be argued that primate cells are held in check by an abundance of low codon usage genes including those likely to lead to rapid growth. Rapid uncontrolled growth often spells disaster in PRI, in contrast to unicellular organisms, and our findings can be rationalized in this way. Alternately, as noted above, low-usage codons may not play the same role in PRI as they do in the other species discussed here.

(e) Codon usage by 'zero class' proteins is more selective than for average proteins

It has been pointed out that proteins in the abundant class have a more restricted codon usage. We have seen that for ECO, YSC and DRO, the zero class also correlates reasonably well with proteins found in the most abundant group. We have engaged in the reciprocal process of selecting the abundant proteins first and then determining their codon usage from the available information in GenBank (results not shown). This reciprocal approach is similar to that taken by Sharp and Li (1987), in which these authors used abundant proteins in ECO and YSC to determine a 'Codon Adaptation Index'. This approach gives very similar results for ECO and YSC, but is more difficult to apply in dealing with organisms that have differentiated cells. Our approach of selecting combinations of excluded codons is more systematic and subject to computer analysis with a minimum of preparation.

Codon usage for zero class proteins is compared with codon usages for all proteins in Table VII. Only data for the major codon families are presented. It can be seen that the most used codons are usually the same in both cases. Exceptions (a total of eight; marked with asterisks) are usually close calls. In many cases for ECO, YSC, and DRO but not PRI, the differences between the most used codons and other codons are more extreme. Indeed, there are more codons in the low-usage group (5% or less) in the highly restricted zero class collection: nine additional in the case of ECO, 14 additional in the case of YSC, eight additional in the case of DRO, but only one additional in the case of PRI. Many of these additional codons in the 5% or less groupings may be low-usage codons in the sense that they may translate more slowly under some or all conditions of growth. This would be a reasonable explanation for their being more scarce in proteins that are synthesized in larger amounts. To decide this, as in all cases, actual measurements of translation rates will have to be made for each of the codons individually.

(f) Choices of low-usage codons are relatively insensitive to gross base composition

To begin a consideration of the origin of varying low-usage codons in different species, we might first examine the relationship between codon usage and base frequencies in reading frames. This information is presented in Table VIII as the ratio of the observed number of codons vs. the expected number of codons calculated from base frequencies within reading frames on the coding strand. The most relevant relationships are for synonym codons.

A ratio greater than one indicates a codon that appears at a higher-than-expected frequency based on the observed base composition of the reading frames. Similarly, a ratio of less than one indicates a codon that appears at lower-than-expected frequency. In almost all cases, the numbers presented in Table VIII reflect the absolute numbers presented in Table I. For example, the ratios of 0.21 and 3.43 for the ECO Leu codons CUA and CUG in Table VIII reflect the much lower (960) and much higher (17477) numbers for these codons in Table I. Thus, the extreme numbers

TABLE VII

Comparison of codon usage[a] for all proteins (T)[b] and for 'zero-class'-proteins (Z)[c] for major codon families

		ECO		YSC		DRO		PRI	
		T	Z	T	Z	T	Z	T	Z
Leu	UUA	11	3	27	17	6	0	6	8
	UUG	11	5	36	75	13	14	11	11
	CUU	10	4	11	2	9	6	11	8
	CUC	10	5	5	0	15	17	21	18
	CUA	3	0	13	6	8	2	7	7
	CUG	55	83	9	1	44	61	45	49
Ile	AUU	47	26	50	51	33	30	33	25
	AUC	46	74*	30	47	53	70	53	58
	AUA	7	0	20	1	14	0	13	18
Val	GUU	29	47*	44	56	19	18	16	19
	GUC	20	9	25	41	26	38	25	18
	GUA	17	24	16	1	9	2	9	4
	GUG	34	19	15	3	46	42	49	59
Ser	UCU	18	36*	31	53	8	12	18	28*
	UCC	17	32	18	32	26	51	24	26
	UCA	12	3	19	6	8	2	13	11
	UCG	14	0	3	1	22	15	6	0
	AGU	13	4	15	4	12	0	13	11
	AGC	26	25	9	4	24	17	25	23
Pro	CCU	15	14	29	14	11	10	27	43*
	CCC	10	0	13	2	36	65	35	40
	CCA	19	14	49	84	23	19	26	17
	CCG	56	72	9	0	30	5	11	0
Thr	ACU	20	38	38	51	16	13	23	32
	ACC	45	52	24	45	42	80	40	35
	ACA	12	3	26	3	17	3	25	32
	ACG	23	7	11	0	25	4	12	0
Ala	GCU	19	40*	44	72	13	27	28	33
	GCC	24	10	24	24	49	68	42	46
	GCA	22	24	24	3	15	4	20	22
	GCG	35	25	8	0	17	3	11	0
Arg	CGU	45	67	17	13	19	44	9	0
	CGC	37	33	4	0	33	50	22	0
	CGA	5	0	5	0	13	0	10	0
	CGG	8	0	2	0	14	0	20	0
	AGA	4	0	54	87	9	0	19	48
	AGG	2	0	17	0	12	6	21	52*
Gly	GGU	38	53*	60	93	22	31	15	16
	GGC	40	42	15	4	43	42	35	40
	GGA	9	3	15	2	28	26	24	38
	GGG	13	3	9	1	7	0	25	8

[a] Numbers are % of listed codons for the same aa. The numbers for each aa add to 100%. Asterisks indicate cases where most-used codon is different in zero class proteins.

[b] Values taken directly from Table V.

[c] 'Zero class proteins' are those proteins that contain no residues encoded by any of the designated eight low-usage codons listed in Table IV.

observed for these codons cannot be explained by gross base compositions of the reading frames. This does not mean that gross base compositions have no influence on differential codon usage. For example, consider the case of the Lys codons AAA and AAG in YSC. Table IB shows that there are more AAA than AAG codons (9104 vs. 8536). Nevertheless, AAA codons are closer to the expected ratio than AAG codons (1.18 vs. 1.66), raising the possibility that the more favored base composition of the AAA codon influences the higher frequency of usage of this codon. YSC probably have more examples like this than the other species we are considering here because the base frequencies for YSC differ most from the equimolar value (cf. top of Table VIII,A through D). Analyses of species with more extreme base compositions would be interesting in this regard. Our main conclusion from Table VIII is that low-usage codon choices are only influenced in a minor way by the gross base composition. Other workers have reported analyses indicating that overall codon usage patterns are influenced by gross base compositions (e.g., Bibb et al., 1984; Osawa and Jukes, 1988); however, these analyses primarily reflect codon usage patterns of average and high-usage codons.

(g) Low codon usage appears to be influenced by dinucleotide usage in some cases

In the previous section we saw that the choice of low-usage codons in the four groups of organisms we chose to consider is not influenced to any appreciable extent by the gross base composition. In this section we consider a related possibility that the choice of low-usage codons is influenced by dinucleotide preferences. It has been argued that dinucleotide preferences govern to a large extent codon choices in eukaryotes (Nussinov, 1981; Alff-Steinberger, 1987). Dinucleotide frequencies for the different species are presented in Table IX, together with the ratio of the observed-to-expected frequencies based on gross base composition. If there is a strong bias against the use of certain dinucleotide sequences it should show up in this ratio. Inspection of Table IX shows seven cases (underlined) where this ratio is 0.73 or less. The UA dinucleotide ratio is low for all species (0.71 ECO; 0.69 YSC; 0.58 DRO; 0.53 PRI), the CG ratio is low for both YSC (0.72) and PRI (0.48) and the AG ratio is low for ECO (0.73). As was seen in Table IV, ECO contains two low-usage codons, AGG and AGA, with the AG sequence, and two low-usage codons, CUA and AUA, with the UA sequence. YSC contains six low-usage codons, CGA, CGG, CGC, CCG, GCG, ACG, with the CG sequence. DRO contains two low-usage codons, AUA and UUA, with the UA sequence. Finally, all eight of the low-usage codons in PRI contain the CG sequence. On the basis of these correlations alone we must consider the proposition that there is some unknown pressure or pressures that cause certain dinucleotide sequences to be underrepresented which in turn results in an underrepresentation of the related codon or codons. However, in the process of attempting such an analysis we must be careful to distinguish cause and effect. Therefore, we must try to determine whether low dinucleotide fre-

TABLE VIII

Ratios of observed number of codons vs. expected number of codons [a]

A. ECO Gross base composition for sum of reading frames: numbers U: 226989 C: 238372 A: 238380 G: 265468

	U	C	A	G	
U	1.47 (Phe)	0.78 (Ser)	1.12 (Tyr)	0.82 (Cys)	U
	1.34 (Phe)	0.87 (Ser)	0.94 (Tyr)	0.88 (Cys)	C
	0.78 (Leu)	0.46 (Ser)	0.14 (STP)	0.05 (STP)	A
	0.75 (Leu)	0.50 (Ser)	0.01 (STP)	0.78 (Trp)	G
C	0.74 (Leu)	0.48 (Pro)	0.80 (His)	1.57 (Arg)	U
	0.69 (Leu)	0.28 (Pro)	0.72 (His)	1.30 (Arg)	C
	0.21 (Leu)	0.55 (Pro)	0.88 (Gln)	0.18 (Arg)	A
	3.43 (Leu)	1.44 (Pro)	1.79 (Gln)	0.25 (Arg)	G
A	2.02 (Ile)	0.76 (Thr)	1.15 (Asn)	0.47 (Ser)	U
	1.90 (Ile)	1.64 (Thr)	1.64 (Asn)	0.90 (Ser)	C
	0.28 (Ile)	0.44 (Thr)	2.52 (Lys)	0.13 (Arg)	A
	1.67 (Met)	0.76 (Thr)	0.72 (Lys)	0.07 (Arg)	G
G	1.38 (Val)	1.13 (Ala)	2.04 (Asp)	1.62 (Gly)	U
	0.80 (Val)	1.40 (Ala)	1.33 (Asp)	1.65 (Gly)	C
	0.76 (Val)	1.25 (Ala)	3.54 (Glu)	0.38 (Gly)	A
	1.40 (Val)	1.79 (Ala)	1.03 (Glu)	0.47 (Gly)	G

B. YSC Gross base composition for sum of reading frames: numbers U: 200820 C: 140270 A: 329877 G: 152556

	U	C	A	G	
U	1.08 (Phe)	1.65 (Ser)	0.87 (Tyr)	0.47 (Cys)	U
	1.34 (Phe)	1.40 (Ser)	0.97 (Tyr)	0.33 (Cys)	C
	0.99 (Leu)	0.90 (Ser)	0.04 (STP)	0.03 (STP)	A
	1.98 (Leu)	0.57 (Ser)	0.02 (STP)	0.84 (Trp)	G
C	0.84 (Leu)	1.92 (Pro)	0.72 (His)	0.66 (Arg)	U
	0.39 (Leu)	0.79 (Pro)	0.70 (His)	0.24 (Arg)	C
	0.59 (Leu)	1.72 (Pro)	1.51 (Gln)	0.17 (Arg)	A
	0.73 (Leu)	0.52 (Pro)	0.80 (Gln)	0.12 (Arg)	G
A	1.26 (Ile)	1.28 (Thr)	1.12 (Asn)	0.63 (Ser)	U
	1.09 (Ile)	1.19 (Thr)	1.32 (Asn)	0.55 (Ser)	C
	0.45 (Ile)	0.78 (Thr)	1.18 (Lys)	1.13 (Arg)	A
	1.16 (Met)	0.51 (Thr)	1.66 (Lys)	0.53 (Arg)	G
G	1.55 (Val)	2.61 (Ala)	2.00 (Asp)	2.81 (Gly)	U
	1.32 (Val)	1.97 (Ala)	1.78 (Asp)	1.02 (Gly)	C
	0.53 (Val)	1.17 (Ala)	2.31 (Glu)	0.83 (Gly)	A
	0.77 (Val)	0.60 (Ala)	1.20 (Glu)	0.54 (Gly)	G

C. DRO Gross base composition for sum of reading frames: numbers U: 75309 C: 103080 A: 98754 G: 103443

	U	C	A	G	
U	1.41 (Phe)	0.55 (Ser)	1.04 (Tyr)	0.50 (Cys)	U
	2.11 (Phe)	1.33 (Ser)	1.45 (Tyr)	0.93 (Cys)	C
	0.46 (Leu)	0.45 (Ser)	0.08 (STP)	0.03 (STP)	A
	1.29 (Leu)	1.11 (Ser)	0.04 (STP)	0.66 (Trp)	G
C	0.65 (Leu)	0.38 (Pro)	0.75 (His)	0.66 (Arg)	U
	0.84 (Leu)	0.93 (Pro)	0.91 (His)	0.88 (Arg)	C
	0.47 (Leu)	0.68 (Pro)	0.79 (Gln)	0.37 (Arg)	A
	2.41 (Leu)	0.77 (Pro)	1.99 (Gln)	0.35 (Arg)	G
A	1.56 (Ile)	0.63 (Thr)	1.65 (Asn)	0.70 (Ser)	U
	1.87 (Ile)	1.26 (Thr)	1.58 (Asn)	1.01 (Ser)	C
	0.56 (Ile)	0.57 (Thr)	0.98 (Lys)	0.29 (Arg)	A
	1.72 (Met)	0.73 (Thr)	2.49 (Lys)	0.33 (Arg)	G
G	0.97 (Val)	0.99 (Ala)	1.95 (Asp)	1.05 (Gly)	U
	1.02 (Val)	1.86 (Ala)	1.38 (Asp)	1.47 (Gly)	C
	0.37 (Val)	0.61 (Ala)	1.08 (Glu)	1.04 (Gly)	A
	1.78 (Val)	0.66 (Ala)	2.45 (Glu)	0.23 (Gly)	G

D. PRI Gross base composition for sum of reading frames: numbers U: 391431 C: 480849 A: 461103 G: 492738

	U	C	A	G	
U	1.52 (Phe)	1.09 (Ser)	1.06 (Tyr)	0.79 (Cys)	U
	1.86 (Phe)	1.18 (Ser)	1.29 (Tyr)	0.92 (Cys)	C
	0.46 (Leu)	0.68 (Ser)	0.05 (STP)	0.09 (STP)	A
	0.88 (Leu)	0.26 (Ser)	0.03 (STP)	0.88 (Trp)	G
C	0.89 (Leu)	1.03 (Pro)	0.55 (His)	0.30 (Arg)	U
	1.35 (Leu)	1.08 (Pro)	0.83 (His)	0.61 (Arg)	C
	0.45 (Leu)	0.85 (Pro)	0.55 (Gln)	0.30 (Arg)	A
	2.83 (Leu)	0.34 (Pro)	1.92 (Gln)	0.54 (Arg)	G
A	1.30 (Ile)	0.90 (Thr)	1.22 (Asn)	0.65 (Ser)	U
	1.69 (Ile)	1.30 (Thr)	1.33 (Asn)	1.03 (Ser)	C
	0.42 (Ile)	0.36 (Thr)	1.39 (Lys)	0.58 (Arg)	A
	1.53 (Met)	0.37 (Thr)	2.05 (Lys)	0.59 (Arg)	G
G	0.83 (Val)	1.27 (Ala)	1.49 (Asp)	0.70 (Gly)	U
	1.05 (Val)	1.56 (Ala)	1.61 (Asp)	1.30 (Gly)	C
	0.41 (Val)	0.78 (Ala)	1.55 (Glu)	0.93 (Gly)	A
	2.00 (Val)	0.38 (Ala)	2.28 (Glu)	0.88 (Gly)	G

[a] Observed number of codons is taken from Table I; expected number of codons is calculated from base frequencies within reading frames on the coding strand, assuming bases were randomly associated to form codons. The 'gross base composition for sum of reading frames' is the actual number of each base counted in coding sequences. For example, in all surveyed reading frames of ECO, there were 226 989 U, 238 372 C, 238 380 A and 265 468 G, for a sum of 969 209 total bases (the sum may not be a multiple of three because we omitted unidentified bases in some sequences). The expected number for a codon is calculated from the probability that any base will occur at a specific position of the codon. In the calculation, the base frequencies are used as the probabilities. In ECO, for example, the probability that a base occurs at a specific position of the codon is $226\,989 \div 969\,209 = 0.23420$ for U; $238\,372 \div 969\,209 = 0.24594$ for C; $238\,380 \div 969\,209 = 0.24595$ for A; $265\,468 \div 969\,209 = 0.27390$ for G. The probability at which a codon is expected to occur is the product of the probabilities of the bases in the codon. The probability of the AGG (Arg) codon of ECO, for example, is calculated as $0.24595 \times 0.27390 \times 0.27390 = 0.01845$. This number means that approx. 1.8% of all codons in ECO would be expected to be AGG (also, another approx. 1.8% would be GAG and still another approx. 1.8% would be GGA). Since a total of 323 059 codons were counted in ECO (Table I), the expected number of AGG codons is therefore $323\,059 \times 0.01845 = 5960$. The actual number of AGG codons counted was 428 (Table I). Thus, the ratio of observed to expected numbers of codons for AGG in ECO is $428 \div 5960 = 0.07$.

quency is the cause of low codon frequency or the other way around.

Let us first consider the AG case for ECO. The two Arg codons AGA and AGG are both low-usage codons in ECO. If we examine the AGX box in which they occur (Table I), we find that this box is shared by two Ser codons and two Arg codons. The Ser codons are well represented compared to other Ser codons. Based on this comparison alone we cannot make a case for saying that these low-usage codons for Arg in ECO are the result of the avoidance of the AG sequence. Other factors must be involved.

Next let us consider the UA case for DRO. Two codons, AUA for Ile and UUA for Leu, have been designated as low-usage codons. If we look at the other codons contain-

TABLE IX

Frequencies of dinucleotides in exons

	ECO		YSC		DRO		PRI	
UU	56879[a]	(1.08)[b]	84409	(1.16)	16404	(1.07)	83987	(1.01)
UC	51707	(0.93)	40645	(1.04)	22614	(1.08)	103345	(1.00)
UA	39532	(0.71)	43974	(0.69)	11023	(0.58)	52063	(0.53)
UG	77688	(1.25)	51154	(1.21)	25971	(1.24)	150276	(1.43)
CU	53439	(0.96)	39507	(1.02)	20708	(0.99)	130750	(1.27)
CC	50942	(0.87)	29845	(1.10)	26884	(0.95)	143974	(1.14)
CA	57384	(0.98)	49444	(1.11)	30346	(1.19)	143161	(1.18)
CG	76386	(1.17)	21381	(0.72)	25081	(0.89)	62330	(0.48)
AU	59307	(1.07)	58264	(0.93)	21265	(1.13)	91864	(0.94)
AC	54289	(0.93)	41907	(0.94)	22519	(0.88)	104960	(0.87)
AA	75435	(1.30)	81823	(1.13)	23885	(1.03)	122009	(1.06)
AG	47506	(0.78)	46775	(0.97)	25589	(1.00)	140078	(1.13)
GU	56386	(0.91)	38151	(0.90)	17688	(0.85)	83357	(0.79)
GC	81433	(1.25)	27873	(0.94)	31082	(1.10)	128562	(0.99)
GA	63548	(0.98)	53207	(1.10)	27902	(1.08)	140390	(1.13)
GG	83482	(0.87)	38017	(1.03)	26586	(0.94)	138971	(1.05)
sum	965333		721376		375606		1820027	

[a] Numbers to the left refer to totals for all reading frames.

[b] Numbers in parentheses refer to ratios of observed over expected where expected is calculated from gross base composition of the reading frames. Ratios of 0.73 or less are underlined. As an example, for the GA dinucleotide in ECO, the expected number would be calculated as follows. The probability that any position will contain a G is 0.27390, or an A is 0.24595 (see footnote to Table VIII for derivation of these values). The probability of finding this dinucleotide is the product of these two individual probabilities ($0.27390 \times 0.24595 = 0.06737$). Thus, out of 965333 dinucleotides, the expected number of GA dinucleotides is 65034 (0.06737×965333). The observed number of GA dinucleotides was 63548. Therefore, the ratio of observed over expected is $63548 \div 65034 = 0.98$.

ing UA (see Table I) we see that these are low, even though they have not been designated as low-usage codons. The other box containing the UA sequence contains the two Tyr codons and two of the stop codons. Since stop codons are always underrepresented this will be a constant factor in reducing the UA frequency. The same is true for other species. If we look at the other species we find that there is a tendency for the NUA codons to be underrepresented in the family boxes (except for UUA and CUA in YSC). Whether the UA sequence is influencing the codon usage or vice versa is hard to say.

To get some further indication on this point we may turn to the dinucleotide frequencies in the intron regions. These are recorded for YSC, DRO, and PRI in Table X for the UA and CG sequences. In all cases the UA dinucleotide ratio is higher for the introns than for the exons. Thus when the coding pressure is lifted, as in the noncoding regions of the introns, the UA sequence gravitates towards the statistically most probable ratio of 1.00. This comparison clearly favors the argument that the low UA dinucleotide frequency in the coding regions is probably caused by the coding pressure.

The CG sequence is associated with six low-usage codons in YSC and all eight in PRI. Since in both of these cases most of the designated low-usage codons contain most of the dinucleotide sequences in question we must carefully scrutinize the possibility that the low value for the CG sequence results from the low frequency of codons using this sequence.

TABLE X

Ratio of observed to expected frequences of UA and CG dinucleotides

		Exons[a]	Introns[b]
YSC	UA	0.69	0.80
	CG	0.72	0.86
DRO	UA	0.58	0.86
	CG	0.88	0.97
PRI	UA	0.53	0.73
	CG	0.48	0.29

[a] Exon ratios are taken from Table IX (b).

[b] Intron ratios are computed from 7582 dinucleotides for YSC, 56497 dinucleotides for DRO, and 686152 dinucleotides for PRI. The calculation is similar to that of the exon ratios (see Table IX, footnote b), however, the gross base composition in this case is taken from the overall base composition, not just the base composition in reading frames.

Inspection of the intron frequencies for YSC and PRI gives different indications for the CG frequency. Thus in YSC the frequency of CG moves closer to the expected (0.72 vs. 0.85), whereas for PRI it moves even further from the expected (0.48 vs. 0.29). Similarly, in YSC the overlapping intercodon C/G dinucleotide in coding sequences also moves closer to the expected, while in PRI the intercodon C/G is found at much lower frequency than expected (data not shown). It would be hazardous to draw any firm conclusions from this comparison. However, taken at face value this would argue that the low codon usage in PRI is dominated by considerations of avoidance of the CG sequence. A possible reason for this avoidance, as already indicated, is that the CG sequence is a site for methylation in PRI (Razin and Riggs, 1980). In YSC, it remains possible that coding pressure influences the dinucleotide frequencies in the coding regions.

(h) Other factors influencing the selection of low-usage codons

Shepherd (1981) noticed that the coding sequences of most reading frames have a bias for the sequence RNY (R = purine; Y = pyrimidine; and N = purine or pyrimidine). We have done extensive analysis of this and the results (unpublished observations) support the view of Shepherd (1981) which was based on a much smaller data base. It is notable that of the designated low-usage codons (Table IV) there is only one example of a RNY sequence. This correlates with the Shepherd view and suggests that the codons selected as low-usage codons in various species may have evolved from less popular sequence arrangements.

(i) Conclusions

(*1*) We have presented new approaches to identify low-usage codons in a reliable fashion. (*2*) We have been able to assign with reasonable confidence (with the possible exception of PRI) up to eight of the lowest-usage codons in several organisms (Table IV). (*3*) Gross base composition and dinucleotide frequencies in general cannot explain choices of low-usage codons; however, dinucleotide usage does show some influence on codon usage in PRI. (*4*) Low-usage codons are clearly avoided in abundant proteins; those proteins containing a high % of low-usage codons are generally cases where an excess of protein could be detrimental. (*5*) In a subsequent paper, we shall propose a model by which low-usage codons may affect translation rates. Also, a more detailed review of our data on codon usage in primates has recently been published (Zhang and Zubay, 1991)

ACKNOWLEDGEMENTS

We acknowledge support by the American Cancer Society, grant No. MV-313A to G.Z., and the National Institutes of Health, grants Nos. GM27711 and 2 SO7 RR05393 to E.G.; the computer resource is supported by the National Institutes of Health, grant No. 2P41 RR00442-20.

REFERENCES

Abate, C., Luk, D., Gentz, R., Rauscher III, F.J. and Curran, T.: Expression and purification of the leucine zipper and DNA-binding domains of Fos and Jun: both Fos and Jun contact DNA directly. Proc. Natl. Acad. Sci. USA 87 (1990) 1032–1036.

Aliff-Steinberger, C.: Codon usage in *Homo sapiens*: evidence for a coding pattern on the noncoding strand and evolutionary implications of dinucleotide discrimination. J. Theor. Biol. 124 (1987) 89–95.

Andersson, S.G.E. and Kurland, C.G.: Codon preferences in free-living microorganisms. Microbiol. Rev. 54 (1990) 198–210.

Bibb, M.J., Findlay, P.R. and Johnson, M.W.: The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. Gene 30 (1984) 157–166.

Bennetzen, J.L. and Hall, B.D.: Codon selection in yeast. J. Biol. Chem. 257 (1982) 3026–3031.

Bonekamp, F., Andersen, H.D., Christensen, T. and Jensen, K.F.: Codon-defined ribosomal pausing in *Escherichia coli* detected by using the *pyrE* attenuator to probe the coupling between transcription and translation. Nucleic Acids Res. 13 (1985) 4113–4123.

Carter, P.W., Bartkus, J.M. and Calvo, J.M.: Transcription attenuation in *Salmonella typhimurium*: the significance of rare leucine codons in the *leu* leader. Proc. Natl. Acad. Sci. USA 83 (1986) 8127–8131.

Chen, G.-F.T. and Inouye, M.: Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes. Nucleic Acids Res. 18 (1990) 1465–1473.

Curran, J.F. and Yarus, M.: Rates of aa-tRNA selection at 29 sense codons in vivo. J. Mol. Biol. 209 (1989) 65–77.

Gouy, M. and Gautier, C.: Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res. 10 (1982) 7055–7074.

Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R.: Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res. 9 (1981) r43–r74.

Grosjean, H.J., De Henau, S. and Crothers, D.M.: On the physical basis for ambiguity in genetic coding interactions. Proc. Natl. Acad. Sci. USA 75 (1978) 610–614.

Grosjean, H. and Fiers, W.: Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene 18 (1982) 199–209.

Gutman, G.A. and Hatfield, G.W.: Nonrandom utilization of codon pairs in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 86 (1989) 3699–3703.

Hoekema, A., Kastelein, R.A., Vasser, M. and De Boer, H.A.: Codon replacement in the *PGK1* gene of *Saccharomyces cerevisiae*: experimental approach to study the role of biased codon usage in gene expression. Mol. Cell. Biol. 7 (1987) 2914–2924.

Ikemura, T.: Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein

genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J. Mol. Biol. 151 (1981) 389–409.

Ikemura, T.: Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. J. Mol. Biol. 158 (1982) 573–597.

Ikemura, T.: Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. 2 (1985) 13–34.

Ingraham, J.L., Maaløe, O. and Neidhardt, F.C.: Growth rate as a variable, In: Growth of the Bacterial Cell. Sinauer Associates, Sunderland, MA, 1983, pp. 267–315.

Konigsberg, W. and Godson, G.N.: Evidence for use of rare codons in the *dnaG* gene and other regulatory genes of *Escherichia coli*. Proc. Natl. Acad. Sci. USA 80 (1983) 687–691.

Kurland, C.G.: Strategies for efficiency and accuracy in gene expression, 1. The major codon preference: a growth optimization strategy. Trends Biochem. Sci. 12 (1987) 126–128.

Murao, K., Tanabe, T., Ishii, F., Namiki, M. and Nishimura, S.: Primary sequence of arginine transfer RNA from *Escherichia coli*. Biochem. Biophys. Res. Commun. 47 (1972) 1332–1337.

Nussinov, R.: Eukaryotic dinucleotide preference rules and their implications for degenerate codon usage. J. Mol. Biol. 149 (1981) 125–131.

Osawa, S. and Jukes, T.H.: Evolution of the genetic code as affected by anticodon content. Trends Genet. 4 (1988) 191–198.

Pongs, O., Kecskemethy, N., Mueller, R., Krah-Jentgens, I., Baumann, A., Kiltz, H.H., Canal, I., Llamazares, S. and Ferrus, A.: *Shaker* encodes a family of putative potassium channel proteins in the nervous system of *Drosophila*. EMBO J. 7 (1988) 1087–1096.

Post, L.E., Strycharz, G.D., Nomura, M., Lewis, H. and Dennis, P.P.: Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase subunit β in *E. coli*. Proc. Natl. Acad. Sci. USA 76 (1979) 1697–1701.

Razin, A. and Riggs, A.D.: DNA methylation and gene function. Science 210 (1980) 604–610.

Robinson, M., Lilley, R., Emtage, J.S., Yarranton, G., Stephens, P., Millican, A., Eaton, M. and Humphreys, G.: Codon usage can affect efficiency of translation of genes in *Escherichia coli*. Nucleic Acids Res. 12 (1984) 6663–6671.

Rojiani, M.V., Jakubowski, H. and Goldman, E.: Relationship between protein synthesis and concentrations of charged and uncharged $tRNA^{Trp}$ in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 87 (1990) 1511–1515.

Sharp, P.M. and Li, W.-H.: Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. Nucleic Acids Res. 14 (1986) 7737–7749.

Sharp, P.M. and Li, W.-H.: The codon adaptation index – a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 15 (1987) 1281–1295.

Sharp, P.M., Cowe, E., Higgins, D.G., Shields, D.C., Wolfe, K.H. and Wright, F.: Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*: a review of the considerable within-species diversity. Nucleic Acids Res. 16 (1988) 8207–8211.

Shepherd, J.C.W.: Method to determine the reading frame of a protein from the purine/pyrimidine genome sequence and its possible evolutionary justification. Proc. Natl. Acad. Sci. USA 78 (1981) 1596–1600.

Sørensen, M.A., Kurland, C.G. and Pedersen, S.: Codon usage determines translation rate in *Escherichia coli*. J. Mol. Biol. 207 (1989) 365–377.

Sprinzl, M., Hartmann, T., Meissner, F., Moll, J. and Vorderwulbecke, T.: Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res. 15 (1987) r53–r188.

Wada, K.-N., Aota, S.-I., Tsuchiya, R., Ishibashi, F., Gojobori, T. and Ikemura, T.: Codon usage tabulated from the GenBank genetic sequence data. Nucleic Acids Res. 18, Suppl. (1990) 2367–2411.

Xu, H. and Boeke, J.D.: Host genes that influence transposition in yeast: the abundance of a rare tRNA regulates Tyl transposition frequency. Proc. Natl. Acad. Sci. USA 87 (1990) 8360–8364.

Yarus, M., Cline, S.W., Wier, P., Breeden, L. and Thompson, R.C.: Actions of the anticodon arm in translation on the phenotypes of RNA mutants. J. Mol. Biol. 192 (1986) 235–255.

Zhang, S. and Zubay, G.: The peculiar nature of codon usage in primates. In: Setlow, J.K. (Ed.), Genetic Engineering, Vol. 13, 1991, pp. 73–113.

FEBS 15257

EXHIBIT C

FEBS Letters 362 (1995) 1–4

Minireview

Differential codon usage: a safeguard against inappropriate expression of specialized genes?

Milton H. Saier, Jr.*

Department of Biology, University of California at San Diego, La Jolla, CA 92093-0116, USA

Received 2 January 1995; revised version received 14 February 1995

Abstract Recent work has suggested that rare codons are sometimes used for the regulation of specialized gene expression in bacteria. Moreover, the cellular levels of certain tRNAs may fluctuate with growth conditions. Evidence implicating such mechanisms in the control of photosynthesis in *Rhodobacter*, solventogenesis in *Clostridium*, sporulation in *Streptomyces*, and fimbrial phase variation in *E. coli* is summarized. It is suggested that such mechanisms will prove applicable to the control of numerous additional specialized functions, and that the empirical tools for testing this possibility are currently available.

Key words: Codon usage; Bacterium; Translation; Gene expression; Sporulation; Photosynthesis; Solventogenesis; Fimbria

1. Introduction

In 1989, Brinkman et al. noted that eukaryotic proteins such as the human tissue type plasminogen activator, prourokinase, and the gp41 protein of HIV, which have a high content of rare codons in their respective genes, are poorly expressed in *E. coli* [1]. Moreover, induction of the expression of any one of these heterologous, plasmid-encoded genes was found to inhibit cell division and cause plasmid instability. Most remarkably, when the bacteria were simultaneously provided with a plasmid bearing the *dnaY* gene, encoding a rare tRNA ($tRNA^{Arg}_{AGA/AGG}$), production of the eukaryotic proteins was increased while plasmid stability and cell viability improved [1].

While these observations were of considerable practical significance to the bioengineer, they foreshadowed observations and experiments that would suggest that the use of rare codons for specialized or differentiation-specific functions in bacteria might provide a general mechanism to ensure proper temporal and spatial expression of the encoding genes. Although this hypothesis is still far from established, work in several laboratories has provided indirect evidence suggesting that rare codon usage is of functional significance in restricting or specifying appropriate gene expression. In this minireview I summarize the evidence concerned with this issue and reiterate the suggestion that the complement of tRNAs found in a particular bacterium under one set of growth conditions may differ from that found under another set of growth conditions.

*Corresponding author. Fax: (1) (619) 534-7108.
E-mail: msaier@ucsd.edu

2. Codon usage and gene expression

All living organisms possess characteristic GC contents and preferred sets of codons used for protein biosynthesis. GC content is a major determinant of codon usage, and codon adaptation indices (CAI values) have proven to provide a reliable, empirically determined estimate of gene expression level for specific groups of organisms [2–4]. tRNA availability during the evolution of an organism may play a significant role in determining its characteristic preferred codon usage. However, genes obtained by horizontal transmission from a phylogenetically divergent organism with GC content and codon usage different from those of the recipient bacterium approach the values characteristic of their newly acquired host only after hundreds of millions of years [5]. This fact suggests, first, that differences in codon usage must have arisen relatively early during prokaryotic evolution, and, second, that the pressure for a newly acquired gene to assume the codon usage of the host organism is minimal. The fact that certain genes exhibiting a relatively high level of specific rare codons can nevertheless be expressed at high levels when cloned behind a strong promoter (see, for example, [6]) has led some investigators to suggest that the use of rare codons does not in fact serve a regulatory function [4]. It should be pointed out in this regard that the inability to demonstrate a regulatory effect with one set of genes expressed under a given set of experimental conditions does not rule out the possibility of an analogous regulatory function for another set of genes expressed preferentially under a different set of conditions. Below I summarize evidence suggesting that various specialized functions, expressed in a variety of bacteria, may be regulated at the translational level by selective use of rare codons in relevant structural genes (see Table 1).

3. Rare codon usage in phototrophic vs. heterotrophic genes in *Rhodobacter*

In 1991, Wu and Saier noted that genes encoding proteins of the photosynthetic apparatuses (reaction center and light harvesting proteins) of the Gram-negative purple bacteria, *Rhodobacter capsulatus* and *R. sphaeroides*, differed in codon usage from that of genes encoding enzymes of the fructose utilization pathway [7]. While most codons occurred with similar frequencies in these two groups of genes, a few were found to predominate, or be present exclusively in one or the other group (see Table 2 for representative examples). Moreover, other genes, such as those involved in nitrogen utilization or carotenoid biosynthesis, that were expressed under both

SSDI 0014-5793(95)00185-9

Table 1
Selective use of rare codons postulated to control specialized functions in bacteria

Function	Organism	Codon	Amino Acid
Photosynthesis	*Rhodobacter capsulatus*	GCU	Ala
		CUC	Leu
Fructose utilization	*Rhodobacter capsulatus*	AAU	Asn
		UGU	Cys
Solventogenesis	*Clostridium acetobutylicum*	ACG	Thr
Arial mycelium development	*Streptomyces coelicolor*	UUA	Leu
Fimbrial production	*Escherichia coli*	UUG	Leu

heterotrophic and phototrophic conditions, exhibited rare codon usage frequencies that were intermediate between those found in the photosynthetic and fructose-catabolic genes (Table 2). These differences were shown to be statistically significant. It was suggested that different tRNA pools were present under phototrophic vs. heterotrophic growth conditions, and that growth conditions might influence the relative rates of transcription of the tRNA genes and their cognate amino acyl tRNA synthetases. Differences in codon usage might generally allow operation of novel post-transcriptional regulatory mechanisms. It seemed reasonable to suppose that charged tRNA availability and codon usage could provide a safeguard against expression of specialized genes under inappropriate conditions [7].

4. Rare codon usage as a potential regulator of solventogenesis in *Clostridium*

Sauer and Dürre noted in 1992 that a mutational defect preceding the gene *thrA* encoding a rare tRNA, $tRNA^{thr}_{ACG}$ in the low GC Gram-positive bacterium, *Clostridium acetobutylicum*, gave rise to the absence of solventogenesis [8]. This strict anaerobe is a spore-forming bacterium that produces acetone and butanol only during a late stage in the growth cycle. The shift to solventogenesis is accompanied by a series of morphological and physiological changes in motility, shape, and granulose content, culminating in endospore formation. Sauer and Dürre noted that the ACG codon is rarely used and is largely restricted to genes either expressed at the end of exponential growth or involved in the inducible uptake or metabolism of minor carbon and nitrogen sources [8]. Because these investigators did not conduct statistical analyses, it was not possible to state that the observed differences in codon usage reflected a unique characteristic of specific groups of genes encoding specialized functions rather than depressed levels of expressivity [4]. Nevertheless, the potential implications of the observations were clear. As in the case of phototrophic vs. heterotrophic gene expression in *Rhodobacter*, codon usage in *Clostridium* may provide a safeguard to insure proper expression of certain stationary phase vs. log phase genes.

5. Codon usage as a determinant of the differentiated state in *Streptomyces*

Species of the high GC Gram-positive genus *Streptomyces* undergo fungal-like differentiation with the sequential formation of vegetative and aerial mycelia [9,10]. The fact that only the latter structures contain spores reflects the spatial and temporal constraints imposed upon the process of terminal differentiation within this genus. The industrial importance of these organisms is related to their capacity to produce an array of antibiotics and useful secondary metabolites during the post-exponential growth phase. Although these strict aerobes have many of the enzymatic attributes of their low GC Gram-positive cousins, their regulatory mechanisms appear to be remarkably different [11–13].

Leskiw et al. [14] and Fernández-Moreno et al. [15] first observed that, in *Streptomyces coelicolor*, a genetic defect in the gene *bldA*, encoding a rare tRNA, $tRNA^{leu}_{UUA}$ [16,17], blocked aerial mycelium formation and prevented efficient phenotypic expression of several genes containing the rare UUA codon. *bldA* mutations (including deletions) did not interfere with vegetative growth but did prevent aerial mycelium formation and antibiotic production (see [18] for a review). It was suggested that this rare codon occurred preferentially in genes concerned with differentiation and antibiotic production as contrasted with those required for vegetative growth.

More recently, evidence was presented suggesting that mature $tRNA^{leu}_{UUA}$ accumulates in ever increasing amounts as *S. coelicolor* cultures age, and that the temporally regulated accumulation of this mature tRNA species correlates with an increase in efficiency of UUA-containing messenger RNA transcription and/or translation ([19]; but see also [20]). It seemed to exert regulatory effects on events occurring during late growth, including morphological differentiation and antibiotic production.

6. Rare codon usage and the control of fimbrial production in *E. coli*

A recently noted example of potential rare codon control of

Table 2
Examples of differential codon usage in photosynthetic versus heterotrophic genes in *Rhodobacter*

	Codon	Amino acid	Fractional codon usage for each amino acid			
			fru	*pho*	*nif*	Total
Elevated codon usage in *pho* genes	GCU	Ala	0.02	0.11	0.03	0.05
	CUC	Leu	0.11	0.35	0.16	0.19
Elevated codon usage in *fru* genes	AAU	Asn	0.44	0.05	0.18	0.18
	UGU	Cys	0.25	0.00	0.13	0.12

Data were taken from [7]. Numerical values provide the fractional codon usage for each of the four amino acids (Ala, Asn, Cys, and Leu) in the four catagories indicated with abbreviations as follows: *fru*, fructose utilization (heterotrophic) genes, total of 1724 codons analyzed for *R. capsulatus*; *pho*, photosynthetic genes, total of 1976 codons analyzed for *R. capsulatus*; *nif*, nitrogen utilization genes (mostly concerned with nitrogen fixation), total of 3824 codons analyzed for *R. capsulatus*; Total, all available sequenced genes for *Rhodobacter* species at the time of analysis [7].

specialized gene expression concerns the production of type 1 fimbriae in the Gram-negative enteric bacterium, *Escherichia coli* strain F18, which is able to colonize the mouse colon [21]. Burghoff et al. [22] isolated a 6.5 kb *E. coli* sequence that enhanced the colonizing ability of strain F18 and simultaneously stimulated synthesis of type 1 fimbriae. The gene responsible for this stimulation proved to be the *leuX* gene, encoding a tRNA specific for the rare leucine codon UUG. This gene is in single copy at 97 min on the *E. coli* chromosome, and the encoded tRNA species (LeuX) is apparently dispensable for growth [23]. No effect on growth rate was observed when *leuX* was mutated [24]. Another $tRNA^{leu}$, LeuZ, specific for the UUA leucine codon, presumably recognizes UUG by 'wobble', and can thereby substitute adequately for LeuX, at least with respect to the expression of genes encoding functions required for vegetative growth.

The mechanism by which *leuX* gene expression influences type 1 fimbrial production is probably complex. The *fimA* gene, encoding the principal type 1 fimbrilin, lacks UUG codons altogether [25]. However, synthesis of type 1 fimbriae is subject to phase variation due to inversion of a 314 bp DNA segment that includes the *fimA* promoter ([26]; but see also [27]). The ratio of the products of two *fim* genes, *fimB* and *fimE*, determine the frequencies of inversion in the two opposing directions with high levels of FimB favoring the 'off' to 'on' transition. Since *fimB* has six UUG codons while *fimE* has only two [28], it has been proposed that LeuX influences type 1 fimbrial production by controlling *fimB* expression more stringently than that of *fimE* [29]. In this regard it is interesting to note that *leuX* expression is apparently regulated by two proteins (of 22 and 26 kDa) encoded by genes adjacent to *leuX*. Deletion analyses have suggested that the 22 kDa protein is a transcriptional activator while the 26 kDa protein is a repressor of *leuX* expression. These proteins may therefore be indirect regulators of type 1 fimbrial phase variation, and consequently of net fimbrial production.

Various *E. coli* strains are collectively capable of producing at least six distinct virulence-related fimbriae, each exhibiting specificity for and mediating adhesion to a specific mammalian cell surface macromolecule [30,31]. Expression of these fimbriae is often subject to phase variation in agreement with the belief that successful colonization of the host depends on the timely expression and subsequent silencing of specific virulence-related genes, depending on the stage of infection. A recent analysis has revealed that the *leuX* gene of uropathogenic *E. coli* strain 536 encompasses one of several sites responsible for genetic instability [32]. Internal to *leuX* is one of two 18-nucleotide direct repeats that serve as functional sites for excision of a 190 kb DNA segment. This segment encodes, among other functions, P-related fimbriae. Excision of this DNA segment silences expression of *leuX* (possibly controlling type 1 fimbrial synthesis, as noted above) as well as expression of the genetic apparatus encoding P-related fimbriae. As bacterial cells lacking 'excess DNA baggage' and incapable of making fimbriae divide with increased growth rates, it may be that timely excision provides the bacterium that has already established itself in the host organism with pathogenic advantage [32]. Based on the proposed regulatory role of rare tRNAs in controlling fimbrial production, we suggest that it was not accidental that tRNA loci have come to serve as sites of virulence-associated DNA insertion/deletion phenomena.

7. Conclusions and perspectives

How important are the postulated regulatory mechanisms giving rise to codon-controlled phenotypic gene expression? Are they generally operative for the control of starvation-induced or stress-related vs. vegetative gene expression in *E. coli* and other bacteria [33]? Do they function to safeguard proper temporal expression of sporulation (*spo*)-specific genes at any one stage or during several different stages in the well-defined programs of differentiation of various *Bacillus* species [34]? Do they play a role in the control of growth phase-specific or condition-selective gene expression, e.g. expression of genes concerned with bioluminescence in *Vibrio* species [35,36], bacteriorhodopsin-mediated photosynthesis in archaebacteria [37], or induction of virulence-specific genes in bacterial pathogens of plants, animals and other bacteria [38–41]?

The first step towards answering these important questions would seem to be to analyze functionally related groups of bacterial genes for statistically significant differences in codon usage, as reported by Wu and Saier [7] for the photosynthetic vs. heterotrophic genes of *Rhodobacter*. A second step would be to measure variations in the cellular concentrations of specific rare tRNA species made under relevant but differing physiological conditions. The third step would be to establish a causal relationship between rare codon occurance, tRNA level and gene expressivity. Such studies may lead to recognition of novel codon usage-mediated mechanisms for ensuring the proper expression of temporally and spatially regulated genes in prokaryotic microorganisms. The relevance of such mechanisms to eukaryotic organisms, including protozoa, fungi, plants and animals, could then be ascertained by the application of straight-forward comparative approaches.

Acknowledgements: I thank Lisa Preble for expert assistance in the preparation of this manuscript. Work in the author's laboratory was supported by Public Health Service Grants 5RO1AI21702 and 2RO1AI14176.

References

[1] Brinkman, U., Mattes, R.E. and Burke, P. (1989) Gene 85, 109–114.
[2] Holm., L. (1986) Nucleic Acids Res. 14, 3075–3087.
[3] Sharp, P.M. and Li, W.H. (1987) Nucleic Acids Res. 15, 1281–1295.
[4] Andersson, S.Y.E. and Kurland, C.G. (1990) Microbiol. Rev. 54, 198–210.
[5] Groisman, E.A., Sturmoski, M.A, Solomon, F.R., Lin, R. and Ochman, H. (1993) Proc. Natl. Acad. Sci. USA 90, 1033–1037.
[6] Robinson, M., Lilley, R., Little, S., Emtage, J.S., Yarranton, G., Stephens, P., Millican, A., Easton, M. and Humphreys, G. (1994) Nucleic Acids Res. 12, 6663–6667.
[7] Wu, L.-F. and Saier Jr., M.H. (1991) Res. Microbiol. 142, 943–949.
[8] Sauer, U. and Dürre, P. (1992) FEMS Microbiol. Lett. 100, 147–154.
[9] Chater, K.F. (1989) in: Genetics and Molecular Biology of Industrial Microorganisms (Hershberger, C.L., Queener, S.W. and Hegeman, G. eds.) pp. 99–107, American Society for Microbiology, Washington, DC.
[10] Chater, K.F. (1993) Annu. Rev. Microbiol. 47, 668–713.
[11] Kwakman, J.H.J.M. and Postma, P.W. (1994) J. Bacteriol. 176, 2694–2698.
[12] Tügemeyer, F., Walkenhorst, J., Cui, X., Reizer, J. and Saier Jr., M.H. (1994) Res. Microbiol. 145, 89–92.

[13] Titgemeyer, F., Walkenhorst, J., Reizer, J., Stuiver, M.H., Cui, X. and Saier Jr., M.H. (1995) Microbiology 141, 51–58.
[14] Leskiw, B.K., Lawlor, E.J., Fernandez-Abalos, J.M. and Chater, K.F. (1991) Proc. Natl. Acad. Sci. USA 88, 2461–2465.
[15] Fernández-Moreno, M.A., Caballero, J.L., Hopwood, D.A. and Malpartida, F. (1991) Cell 66, 769–780.
[16] Bibb, M.J., Findlay, P.R., and Johnson, M.W. (1984) Gene 30, 157–166.
[17] Wright, F. and Bibb, M.J. (1992) Gene 113, 55–65.
[18] Leskiw., B.K., Bibb, M.J. and Chater, K.F. (1991) Mol. Microbiol. 5, 2861–2867.
[19] Leskiw, B.K., Mah, R., Lawlor, E.J. and Chater, K.F. (1993) J. Bacteriol. 175, 1995–2005.
[20] Gramajo, H.C., Takano, E. and Bibb, M.J. (1993) Mol. Microbiol. 7, 837–845.
[21] Cohen, P.S., Rossoll, R., Cabelli, V.J., Yang, S.L. and Laux, D.C. (1983) Infect Immun. 40, 62–69.
[22] Burghoff, R.L., Pallesen, L., Krogfelt, K.A., Newman, J.V., Richardson, M., Bliss, J.L., Laux, D.C. and Cohen, P.C. (1993) Infect. Immun. 61, 1293–1300.
[23] Yoshimura, M., Inokuchi, H. and Ozeki, H. (1984) J. Mol. Biol. 177, 627–644.
[24] Komine, Y., Adachi, T., Inokuchi, H. and Ozeki, H. (1990) J. Mol. Biol. 212, 579–598.
[25] Klemm, P. (1984) Eur. J. Biochem. 143, 395–399.
[26] Abraham, J.M., Freitag, C.S., Clements, J.R. and Eisenstein, B.I. (1989) Proc. Natl. Acad. Sci. USA 82, 5724–5727.
[27] McClain, M., Blomfield, I.C., Eberhardt, K.J. and Eisenstein, B.I. (1993) J. Bacteriol. 175, 4335–4344.
[28] Klemm, P. (1986) EMBO J. 5, 1389–1393.
[29] Newman, J.V., Burghoff, R.L., Pallesen, L., Krogfelt, K.A., Kristensen, C.S., Laux, D.C. and Cohen, P.S. (1994) FEMS Microbiol. Lett. 122, 281–287.
[30] Van Rosmalen, M. and Saier Jr., M.H. (1993) Res. Microbiol. 144, 507–527.
[31] Klemm, P. (1994) Fimbriae: Adhesion, Genetics, Biogenesis, and Vaccines, CRC Press, Boca Raton, FL.
[32] Blum, G., Ott, M., Lischewski, A., Ritter, A., Imrich, H., Tschäpe, H. and Hacker, J. (1994) Infect. Immun. 62, 606–614.
[33] Matin, A. (1990) FEMS Microbiol. Ecol. 74, 185–196.
[34] Losick, R. and Stragier, P. (1992) Nature 355, 601–604.
[35] O'Toane, D.J. and Prasher, D.C. (1992) Mol. Microbiol. 6, 443–449.
[36] Meighen, E.A. (1993) FASEB J. 7, 1016–1022.
[37] Kuan, G. and Saier Jr., M.H. (1994) Res. Microbiol. 145, 273–285.
[38] Thomashow, M.F. and Cotter, T.W. (1992) J. Bacteriol. 174, 5767–5771.
[39] Gross, R. (1993) FEMS Microbiol. Rev. 104, 301–326.
[40] Briggs, S.P. and Johal, G.S. (1994) Trends Genet. 10, 12–16.
[41] Saier Jr., M.H. (1994) FEBS Lett. 337, 14–17.

EXHIBIT D

Volume 12, Supplement

Compilation of sequences of tRNA genes

Mathias Sprinzl* and Dieter H.Gauss+

*Department of Biochemistry, University of Bayreuth, P.O.Box 3008, D-8580 Bayreuth, and
+Department of Chemistry, Max-Planck-Institut für experimentelle Medizin, Hermann-Rein-Strasse 3, D-3400 Göttingen, FRG

INTRODUCTION

In the compilation of tRNA genes the sequences have been aligned and displayed as has been done in the case of the tRNA sequences (Fig. 1 in accompanying compilation). The nucleotides preceding nucleotide residue 1 and the nucleotides following nucleotide residue 73 and the intervening sequences (see footnotes) have been excluded from the compilation. The compilation is deposited with the Nucleotide Sequence Data Library of EMBL, Heidelberg, and available there on magnetic tape upon request. The compilers would welcome any information regarding missing material or erroneous presentation.

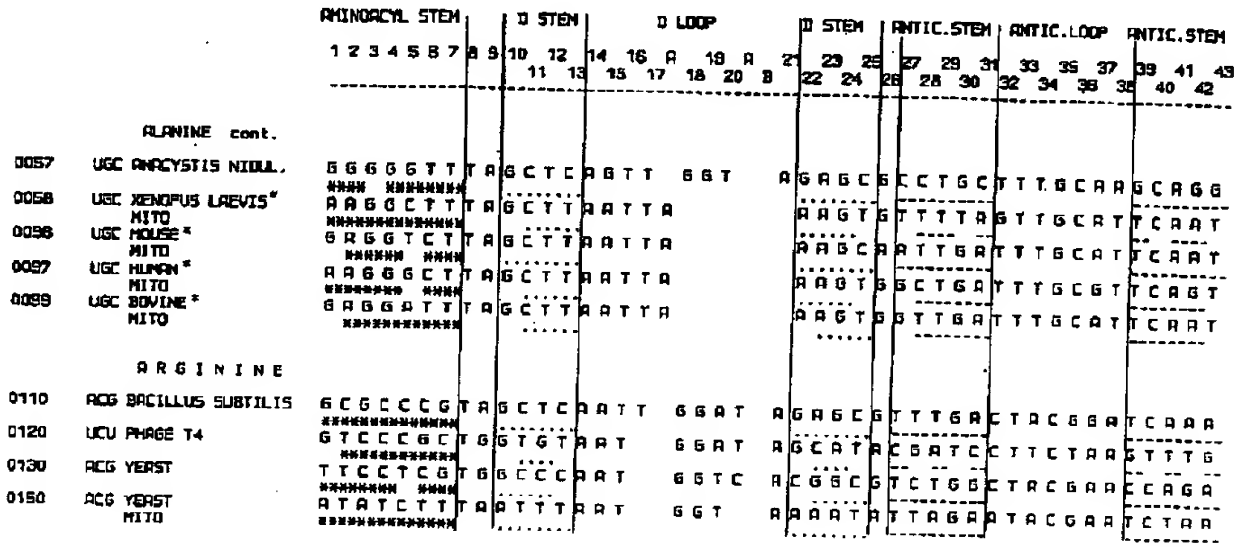
Acknowledgements: We thank Hedwig Geißler, Thomas Hartmann, Werner Heubeck and Rudolf Jung for skillfull cooperation and the Fonds der Chemischen Industrie for financial support.

			AMINOACYL STEM								D STEM		D LOOP							D STEM		ANTIC.STEM			ANTIC.LOOP				ANTIC.STEM		
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	A	18	B	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
	ALANINE																																													
0005	UGC	ETHANOCOCCUS VANI.	G	G	G	C	C	C	G	T	A	G	C	T	C	A	G	T	T				G	G	G		A	G	A	G	C	G	C	T	G	C	C	C	T	T	G	C	A	A	G	G	C	A	G
0010	UGC	BACILLUS SUBTILIS	G	G	G	G	C	C	T	T	A	G	C	T	C	A	G	C	T				G	G	G		A	G	A	G	C	G	C	C	T	G	C	T	T	T	G	C	A	C	G	C	A	G	G
0025	UGC	NEUROSPORA CRASSA MITO	G	G	G	G	G	T	A	T	A	G	T	A	T	A	A	T	T				G	G	T		A	G	T	A	C	A	G	C	A	A	T	C	T	T	G	C	T	C	A	T	T	G	C
0050	AGC	BOMBYX MORI	G	G	G	G	G	C	G	T	A	G	C	T	C	A	G	A	T				G	G	T		A	G	A	G	C	G	C	T	C	G	C	T	T	A	G	C	A	T	G	C	G	A	G
0051	UGC	ASPERGILLUS NIDUL. MITO	G	G	G	G	C	T	A	T	A	G	T	T	T	A	A	C	T				G	G	T		A	A	A	A	C	G	G	C	G	A	T	T	T	G	C	A	T	A	T	C	G	T
0052	UGC	EUGLENA GRACILIS* CHLORO	G	G	G	G	G	T	A	T	A	G	C	T	C	A	G	T	T				G	G	T		A	G	A	G	C	G	C	T	G	C	C	T	T	T	G	C	A	A	G	G	C	A	G
0053	UGC	MAIZE CHLORO	G	G	G	G	A	T	A	T	A	G	C	T	C	A	G	T	T				G	G	T		A	G	A	G	C	T	C	C	G	C	T	C	T	T	G	C	"A	T	G	G	C	G	G
0054	UGC	YEAST MITO	G	G	G	G	T	T	A	T	A	G	T	T	A	A	A	T	T	T	T	G	G	T		G	A	A	A	C	G	A	C	T	G	C	G	T	T	G	C	A	T	G	C	A	T	T
0055	UGC	RAT * MITO	G	A	G	G	A	T	T	T	A	G	C	T	T	A	A	T	T	A								A	A	G	C	A	G	T	T	G	A	T	T	T	G	C	A	T	T	T	A	A	C
0056	UGC	TOBACCO CHLORO	G	G	G	G	A	T	A	T	A	G	C	T	C	A	G	T	T				G	G	T		A	G	A	G	C	T	C	C	G	C	T	C	T	T	G	C	A	"C	G	G	C	G	G

0005 D.JARSCH, A.BOECK(1983) NUCL.ACIDS RES. 11, 7537-7544
0010 K.LINDAHL, E.LUND, J.E.DAHLBERG (1982) NUCL. ACIDS RES. 10, 1607-1624
C.J.GREEN, B.S.VOLD (1983) NUCL. ACIDS RES. 11, 5763-5774
E.F.HAMROUSER, J.N.HANSEN (1983) J. BIOL. CHEM. 258, 291-298
0025 M.R.GREEN, R.F.DRIAN, R.A.GOEWERT, R.A.COLLINS (1981) J. BIOL. CHEM. 256, 2027-2034
0050 O.HAGENBUECHLE, D.LARSON, E.I.HALL, K.U.SPRAGUE (1979) CELL 18, 1217-1229
R.L.GARBER,L.P.GAGE(1979) CELL 18, 817-828
0051 W.G.ADECHEL, C.M. LAZARUS, N.BASAK, H.KUENTZEL (1981) CELL 23, 625-633
0052 L.GRAF, H.KOESSEL, E.STUTZ (1980) NATURE 286, 908-910;
E.A.ORDIZO, JR, K.E.BUSHNELL, J.R.ODOM, R.B.HALLICK (1980) J. BIOL. CHEM. 255, 10997-11003
0053 U.KOCH, K.EDWARDS, H.KOESSEL (1981) CELL 25, 203-213
0054 S.D.BONITZ, A.TZAGOLOFF (1980) J. BIOL. CHEM. 255, 9075-9081
0055 P.CANTATORE, C.DE BENEDETTO, G.GADALETA, R.GALLERANI, A.D.KROON, M.HOLTROP, C.LAMAVE, G.PEPE, C.QUAGLIARIELLO, D.SACCONE, E.SBISA (1982) NUCL. ACIDS RES. 10, 3279-3291
R.GROSSKOPF, B.FELDMANN (1981) CURR. GEN. 4, 191-196
0056 F.TAKAIWA, M.SUGIURA (1982) NUCL. ACIDS RES. 10, 3665-3676

		EXTRA ARM														TP STEM			TF LOOP				TF STEM			AMINOACYL STEM					
		45	47	B	D	F	H	J	L	N	P	48	51	53	55	57	59	61	63	65	67	69	71	73	75
		44	46	A	C	E	G	I	K	M	O	49	50	52	54	58	58	60	62	64	66	68	70	72	74	76
	ALANINE																									
0005		A	G	G	C							C	G	T	G	G	G	T	T	C	A	A	A	T	C	C	C	G	C	C	G	G	G	T	C	C	A
0010		A	G	G	T							C	A	G	C	G	G	T	T	C	G	A	T	C	C	C	G	C	T	A	G	G	C	T	C	C	A
0025		T	T	G								T	C	A	A	G	G	T	T	C	A	A	A	T	C	C	T	T	G	T	A	T	C	T	C	C	A
0050		A	G	G	T							A	C	C	G	G	G	A	T	C	G	A	T	A	C	C	C	G	G	C	G	C	T	C	C	A
0051		T	A	T								T	T	C	A	G	G	A	T	C	G	A	G	T	C	C	T	G	A	T	A	A	C	T	C	C	A
0052		A	T	G	T							C	A	G	C	G	G	T	T	C	G	A	G	T	C	C	G	C	T	T	A	T	C	T	C	C	A
0053		A	T	G	T							C	A	G	C	G	G	T	T	C	G	A	G	T	C	C	G	C	T	T	A	T	C	T	C	C	A
0054		T	A	A								T	A	T	G	A	G	T	T	C	A	A	G	T	C	T	C	A	T	T	A	A	C	T	C	C	A
0055		A	G	A								T	G	T	A	A	G	A	T	A	T	A	G	T	C	T	T	A	C	A	G	T	C	C	T	T	A
0056		A	T	G	T							C	A	G	C	G	G	T	T	C	G	A	G	T	C	C	G	C	T	T	A	T	C	T	C	C	A

0052/0 ORITION 1 TO 49 WAS SEQUENCED BY T.MIYATA,R.KIKUNO,Y.OHSHIMA 1982)NUCL.ACIDS RES. 10, 1771-1780
0053/36 AFTER RESIDUE 36.37 OR 36 INTERVENING SEQUENCE
0053/0 ALIGNMENT IS ARBITRARY
0056/37 INTERVENING SEQUENCE AFTER THIS POSITION



0057 S.E.WILLIAMSON, W.F.DOOLITTLE (1983) NUCL. ACIDS RES. 11, 225-235

0058 J.F.H.WONG, S.P.AA, A.E.WILSON, B.A.ROE (1983) NUCL. ACIDS. RES. 11, 4977-5004

0096 A.J.BIBB, R.A.VAN ETTEN, C.T.WRIGHT, M.W.WALBERG, D.A.CLAYTON (1981) CELL 26, 167-180

0097 S.ANDERSON, A.T.BANKIER, B.G.BARRELL, M.H.L.DEBRUIJN, A.R.COULSON J.DROUIN, I.C.EPERON, D.P.NIERLICH, B.A.ROE, F.SANGER, P.H.SCHREIER, A.J.H.SMITH, R.STADEN, I.B.YOUNG (1981) NATURE 290, 457-465

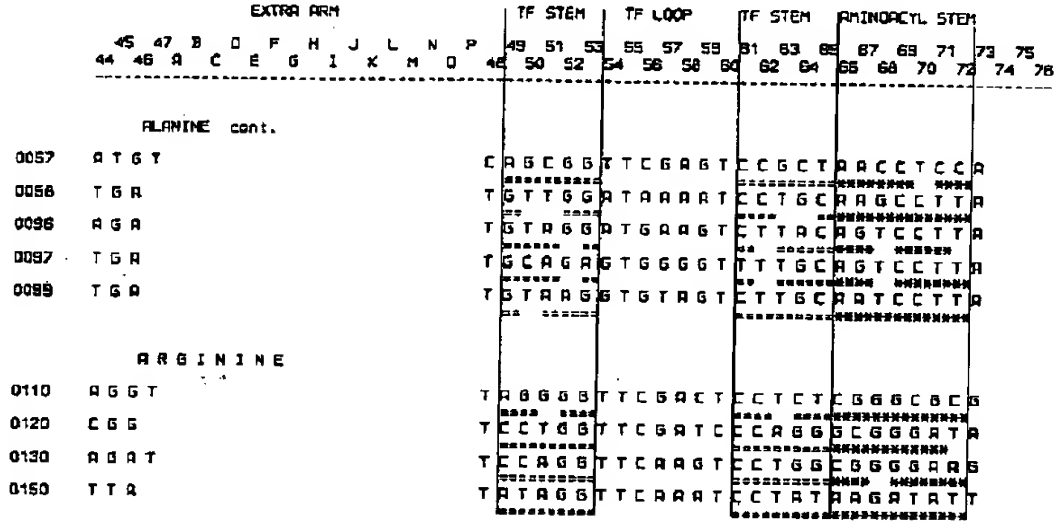
0099 B.ANDERSON, A.N.L.DEBRUIJN, A.R.COULSON, I.C.EPERON, F.SANGER, I.G.YOUNG (1982) J.MOL.BIOL. 156, 683-717

0110 C.J.GREEN, B.S.VOLD (1983) NUCL. ACIDS RES. 11, 5763-5774
C.f. HAUENSTEIN, J.N. HANSEN (1983) J. BIOL. CHEM. 258, 291-295

0120 S.P.MAZZARA, B.PLUMMETT, M.H.MCCLAIR (1981) PROC. NATL. ACAD. SCI. 78, 689-692

0130 R.E.BAKER, A.EIGEL, D.VOEGTEL, H.FELDMANN (1982) EMBO JOURNAL 1, 291-295

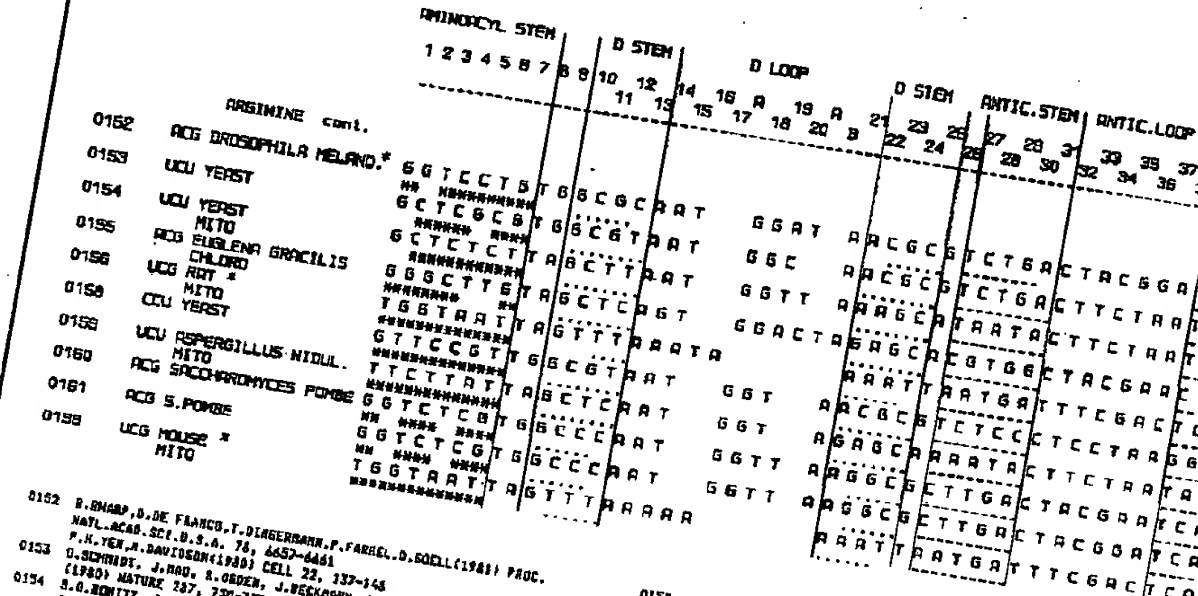
0150 M.C.MARTIN, D.MILLER, J.HARTLEY, P.CUNNINGHAM, J.E.DOWNLSON (1980) CELL 19, 339-343
S.G.BONITZ, A.TZAGOLOFF (1980) J.BIOL.CHEM. 255, 9075-9081
D.L.MILLER, M.C.MARTIN (1981) CURR. GENETICS 4, 135-141



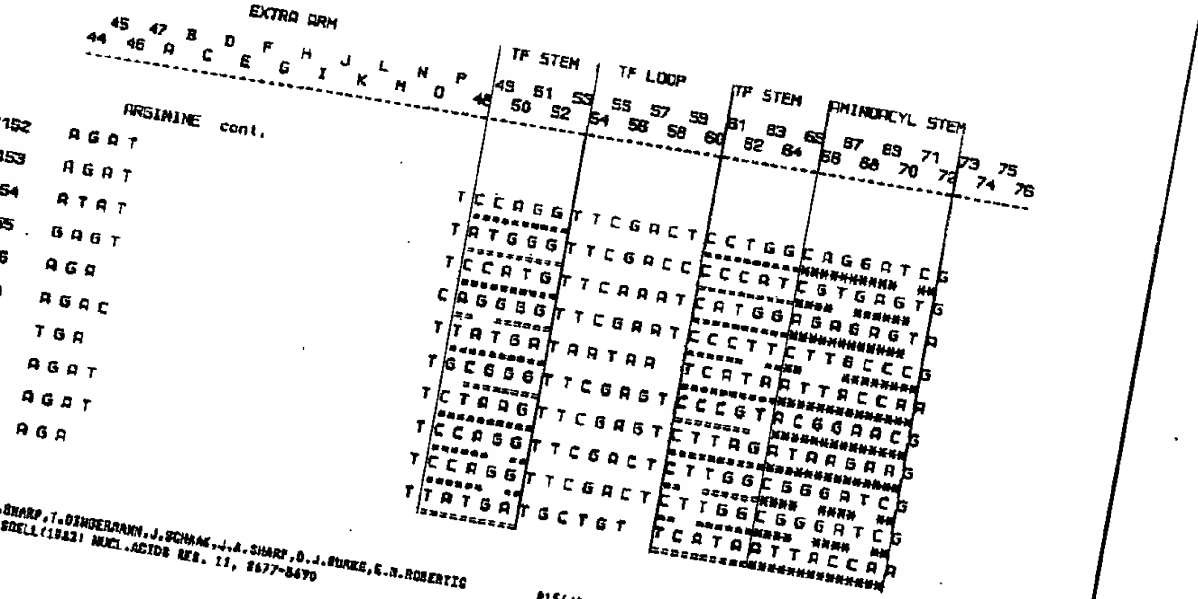
0058/0 ALIGNMENT IS ARBITRARY
0096/0 ALIGNMENT IS ARBITRARY

0097/0 ALIGNMENT IS ARBITRARY
0099/0 ALIGNMENT IS ARBITRARY

819-772-0061

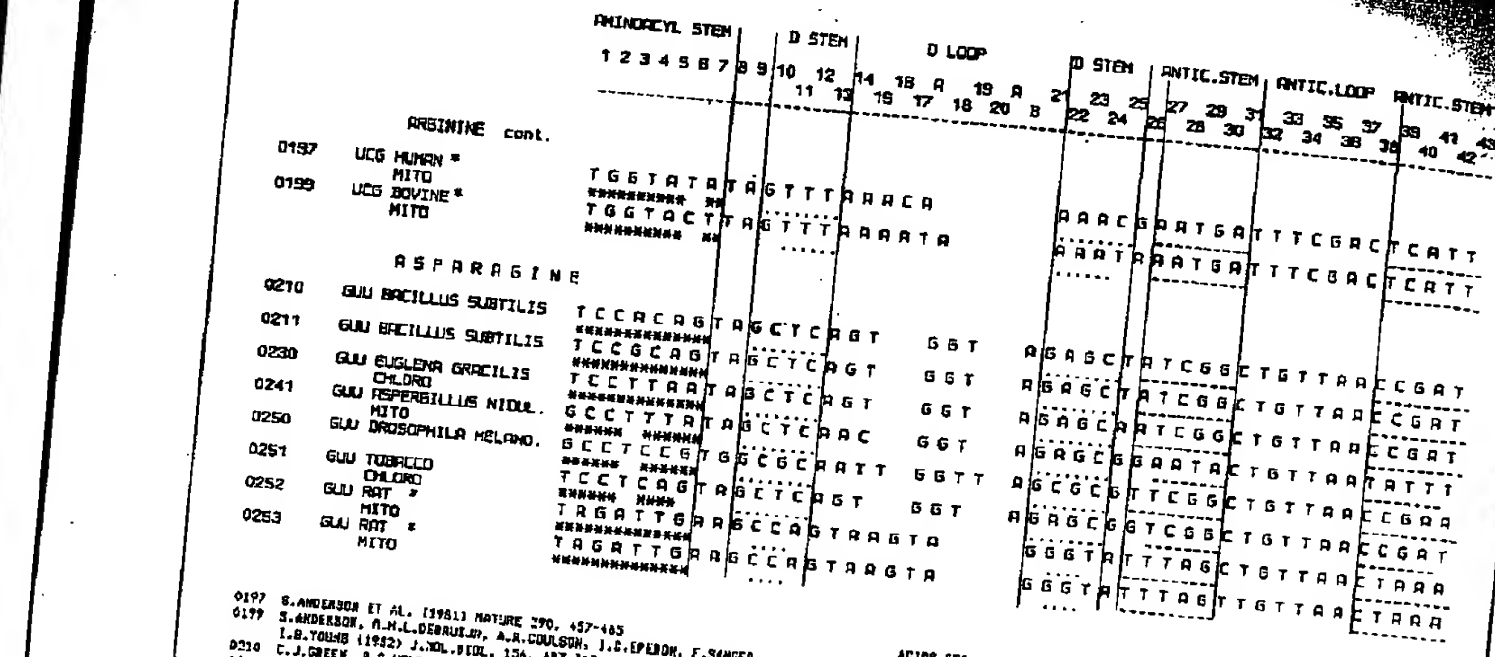


0152 B.SHARP,D.DE FRANCO,T.DINGERMANN,P.FARRELL,D.SOELL(1981) PROC. NATL.ACAD.SCI.U.S.A. 78, 6652-6661
P.K.YEN,R.DAVIDSON(1980) CELL 22, 137-145
0153 U.SCHMIDT, J.RAO, R.GROEBE, J.BECKMANN, H.SAKANO, J.ABELSON, D. SOELL, (1980) NATURE 287, 730-732
0154 S.G.BONITZ, A.IZAGOLOFF (1980) J. BIOL. CHEM. 255, 9075-9081
0155 E.M.DROZCO, JR., R.B. HALLICK (1982) CURR. GEN. 4, 135-143
0156 H.CANTATORE, C.DE BENEDETTO, S.GADALETA, R.GALLERANI, A.M.KROON, D.L.KELLER, R.C.BARTIS (1981) CURR. GEN. 4, 133-143
0158 J.DAWTHER, E.R.DERGOBERTIS, P.PHILIPPSEN (1983) EMBO-J. 2, 583-591
0159 R.METZKER, R.G.IQECHEL, R.BAZAR, H.KUENTZEL (1982) NUCL. ACIDS RES. 10, 4783-4794
E.BRIGL,[.A.BROWN,R.B.VARING,C.SCAZZOCCHIO,R.W.DAVIES(1982) NUCL. ACIDS RES. 10, 3531-3539
0160 V.DAMULIN,J.TAO,B.APPEL,R.SUMNER-SMITH,F.YAMAO,D.SOELL(1983) NUCL. ACIDS RES. 11, 8537-8544
0161 V.DAMULIN,J.RAO,B.APPEL,D.SUMNER-SMITH,F.YAMAO,D.SOELL(1983) NUCL.A RES. 11, 8537-8549
0196 R.J.REID, R.A.VAN ETTEN, C.T.WRIGHT, R.W.WALBERG, D.A.CLAYTON (1982) CELL 28, 167-180

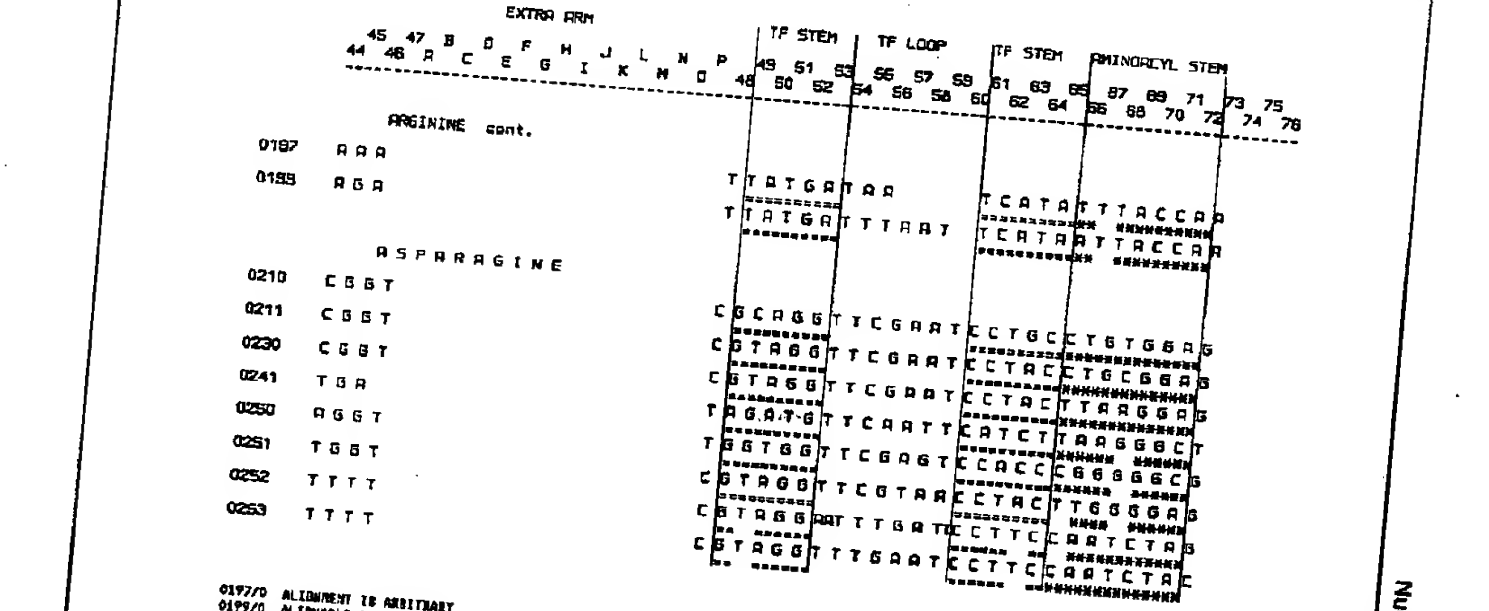


0152/0 B.SHARP,T.DINGERMANN,J.SCHAK,J.A.SHARP,D.J.BURKE,G.R.ROBERTS D.SOELL(1982) NUCL.ACIDS RES. 11, 8677-8690

0156/0 ALIGNMENT IS ARBITRARY
0196/0 ALIGNMENT IS ARBITRARY



0197 S.ANDERSON ET AL. (1981) NATURE 290, 457-465
0199 S.ANDERSON, A.H.L.DEBRUIJN, A.R.COULSON, I.G.EPERON, F.SANGER, I.B.YOUNG (1982) J.MOL.BIOL. 156, 683-717
0210 C.J.GREEN, B.S.VOLD (1983) NUCL. ACIDS RES. 11, 5763-5774
0211 E.F.VANDOOREN, J.H.HANSEN (1983) J. BIOL. CHEM. 258, 291-298
0230 E.R.DROZCO, JR, R.B. HALLICK (1982) J. BIOL. CHEM. 257, 3263-3269
0241 R.HEIZER, H.G.KOECHEL, H.BASAK, H.ZUENTZEL (1982) NUCL. ACIDS RES. 10, 4783-4794
E.GRISI,T.A.BROWN,A.B.HARING,C.SCAZZOCCHIO,R.W.DAVIES(1982) NUCL. ACIDS RES. 10, 3331-3339
0250 B.HOVEMAN, G.SHARP, H.YAMADA, G.SOELL (1980) CELL 19, 889-895
0251 A.KATO, H.SHIMADA, T.KUSUDA, A.SUGIURA (1981) NUCL. ACIDS RES. 9, 5601-5607
0252 F.CANTATORE, C.DE BENEDETTO, S.BADALETA, R.GALLERANI, A.M.ROON, R.HOLTROP, C.LANAVE, G.PEPE, C.DURBOLCARIELLO, C.SACCONE, E.SBISA (1982) NUCL. ACIDS RES. 10, 3279-3289
0253 R.GROSZKOPF, H.FELDMANN (1981) CURR. GEN. 4, 191-196



0197/0 ALIGNMENT IS ARBITRARY
0199/0 ALIGNMENT IS ARBITRARY
0252/0 ALIGNMENT IS ARBITRARY
0253/0 ALIGNMENT IS ARBITRARY

			AMINOACYL STEM								D STEM				D LOOP							D STEM			ANTIC.STEM				ANTIC.LOOP						ANTIC.STEM			
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	A	18	19	B	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
	ASPARTIC ACID cont.																																												
0352	GUC	RAT	T	C	C	T	C	G	T	A	G	T	A	T	A	G	T		G	G	T	G	A	G	T	A	T	C	C	C	G	C	C	T	G	T	C	A	C	G	C	G	G	G
0353	GUC	YEAST	T	C	C	G	T	G	A	T	A	G	T	T	T	A	A	T		G	G	T	C	A	G	A	A	T	G	G	G	C	G	C	T	T	G	T	C	G	C	G	T	G	C	C
0354	GUC	YEAST MITO	G	G	A	T	C	T	G	T	A	G	C	T	T	A	A	T		A	G	T		A	A	A	G	T	A	C	C	A	T	T	T	G	T	C	A	T	A	A	T	G	G
0355	GUC	RAT * MITO	G	A	G	A	T	A	T	T	A	G	T	A	A	A	A	T	A	A						T	T	A	C	A	T	A	A	C	C	T	T	G	T	C	A	A	G	G	T	T	A
0357	GUC	SACCHAROMYCES POMBE	T	C	T	C	C	T	T	T	A	G	T	A	T	A	G	G			G	G	T		A	G	T	A	C	A	C	A	A	G	C	C	T	G	T	C	A	C	G	C	T	T	G
0358	GUC	SPINACH CHLORO	G	G	G	A	T	T	G	T	A	G	T	T	C	A	A	T	T		G	G	T	T	A	G	A	G	C	A	C	C	G	C	C	C	T	G	T	C	A	A	G	G	C	G	G
0365	GUC	DROSOPHILA MELANO.* MITO	A	A	A	A	A	A	T	T	A	G	T	T	A	A	A	A	T	C						A	T	A	A	C	A	T	T	A	G	T	A	T	G	T	C	A	A	A	C	T	A	A
0366	GUC	DROSOPHILA YAKUBA * MITO	A	A	A	A	A	A	T	T	A	G	T	T	A	A	A	T	T	A	T	A				T	A	A	C	A	T	T	A	G	T	A	T	G	T	C	A	A	A	C	T	A	A
0396	GTC	MOUSE	T	C	C	T	C	G	T	T	A	G	T	A	T	A	G	T		G	G	T	G	A	G	T	A	T	C	C	C	G	C	C	T	G	T	C	A	C	G	C	G	G	G
0396	GUC	MOUSE * MITO	A	A	G	A	T	A	T	T	A	G	T	A	A	A	A	T	C	A	A					T	T	A	C	A	T	A	A	C	C	T	T	G	T	C	A	A	A	G	T	T	A

0352 T.SEKIYA, Y.KUCHINO, S.NISHIMURA (1981) NUCL. ACIDS RES. 9,2239-2250; K.SHIBUYA, S.NOGUCHI, S.NISHIMURA, T.SEKIYA (1982) NUCL. ACIDS RES. 10, 4441-4448

0353 G.ROMBIDT, J.HAO, R.GOOEN, J.BECKMANN, H.SAKANO, J.ABELSON, D.SOELL (1980) NATURE 287, 750-752
J.GAFNER,E.N.DE ROBERTIS,P.PHILIPPSE(1983) EMBO J. 2, 583-591

0354 S.B.SONITZ, A.TZAGOLOFF (1980) J. BIOL. CHEM. 255,9075-9081
D.L.MILLER, D.R.MAJARIAN, J.R.FOLSE, M.C.MARTIN (1981) J.BIOL.CHEM. 256, 9774-9777
D.L.MILLER, M.C.MARTIN (1981) CURR. GEN. 4, 135-143

0355 P.CANTATORE, C.DE BENEDETTO, G.GADALETA, R.BALLERANI, A.M.BROON, R.HOLTROP, C.LANAVE, G.PEPE, C.QUAGLIARIELLO, C.SACCONE, E.SBISA (1982) NUCL. ACIDS RES. 10, 3279-3289

0357 J.RAO, B.APPEL, J.SCHRAN, S.SHARP, H.YAMADA, D.SOELL (1982) NUCL. ACIDS RES. 10, 487-500

9357 E.HOLSCHUH,W.STOTTORLEY,P.R.WHITFELD(1983) NUCL.ACIDS RES. 11, 8547-8554

0358 M.M.C.DE BRUIJN(1983) NATURE 304, 234-241

0364 D.C.CLARY, D.R.WOLSTENHOLME (1983) NUCL. ACIDS RES. 11, 4211-4227

0395 J.E.LOONEY,J.D.MARSH(1953) NUCL.ACIDS RES. 11, 8751-8775

0396 M.J.GIBB, R.A.VAN ETTEN, C.T.WRIGHT, M.M.WALBERG, D.A.CLAYTON (1981) CELL 24, 167-130

			EXTRA ARM																							TF STEM			TF LOOP			TF STEM			AMINOACYL STEM										
			45	47	B	D	F	H	J	L	N	P	48	51	53	55	57	59	61	63	65	67	69	71	73	75																				
			44	46	A	C	E	G	I	K	M	O	49	50	52	54	58	58	60	62	64	66	68	70	72	74	76																				
	ASPARTIC ACID cont.																																												
0352			R	G	A								C	C	G	G	G	G	T	T	C	G	A	T	T	C	C	C	C	G	A	C	G	G	G	G	A	G*							
0353			A	G	A								T	C	G	G	G	G	T	T	C	A	A	T	T	C	C	C	C	G	T	C	G	C	G	G	A	G							
0354			A	G	G	A							T	G	T	C	A	G	T	G	C	A	A	A	T	C	T	G	A	T	T	A	G	A	T	T	C	G							
0358			A	G	T								T	A	T	A	G	A	C	C*	T	A	A	A		T	C	T	A	T	A	T	A	T	C	T	T	A							
0357			C	A	G								C	C	C	G	G	G	T	T	C	G	A	A	T	C	C	C	C	G	G	A	G	G	G	A	G	A	G						
0359			A	A	G	C							T	G	C	G	G	G	T	T	C	G	A	G	C	C	C	C	G	T	C	A	G	T	C	C	C	G							
0365			A	A	T								T	A	T	T	A	A	A	T	A	A				T	T	A	A	T	A	T	T	T	T	T	T	A							
0366			A	A	T								T	A	T	T	A	A	A	T	T	A				T	A	A	T	A	T	T	T	T	T	T	A							
0395			A	G	A								C	C	G	G	G	G	T	T	C	G	A	T	T	C	C	C	C	G	A	C	G	G	G	G	A	G							
0396			A	A	T								T	A	T	A	G	A	T	C	A	A	T	A	A	T	C	T	A	T	A	T	A	T	C	T	T	A							

0352/73 IS REPORTED TO BE C IN D.R.MAKOWSKI,R.A.HAAS,H.P.GOLAN,D.BRUNBERGER, (1983) NUCL.ACIDS RES. 11, 8609-8624

0355/0 ALIGNMENT IS ARBITRARY

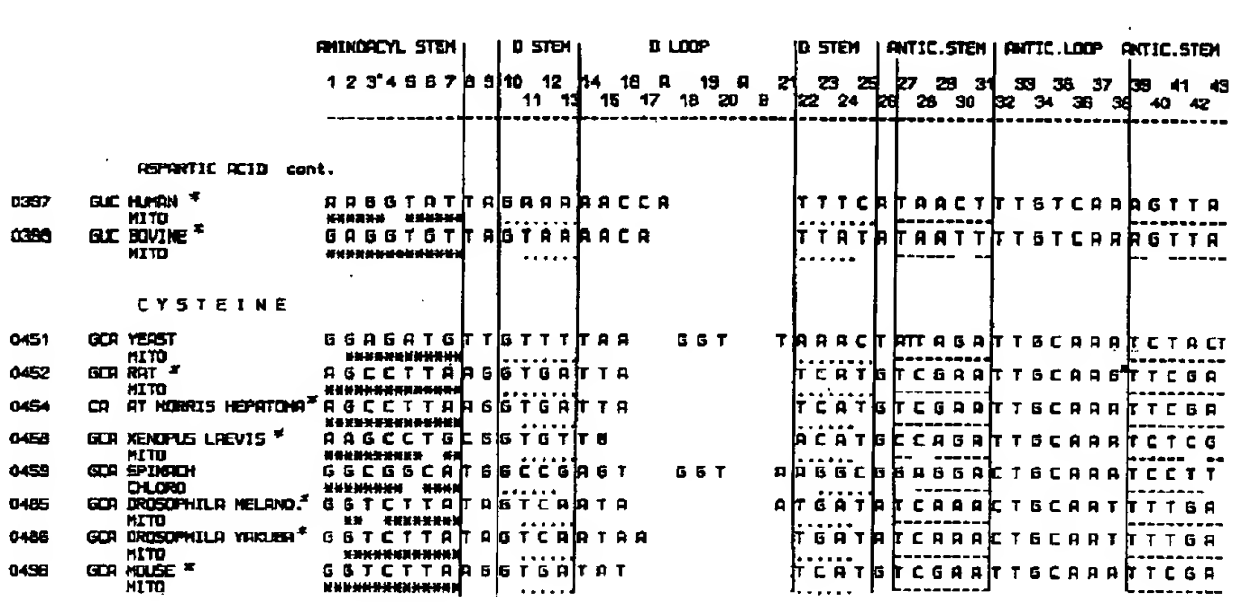
0355/55 IS REPORTED TO BE C IN R.GROSSKOPF, H.FELDMANN (1981) CURR. GEN. 4, 151-158

0365/0 ALIGNMENT IS ARBITRARY

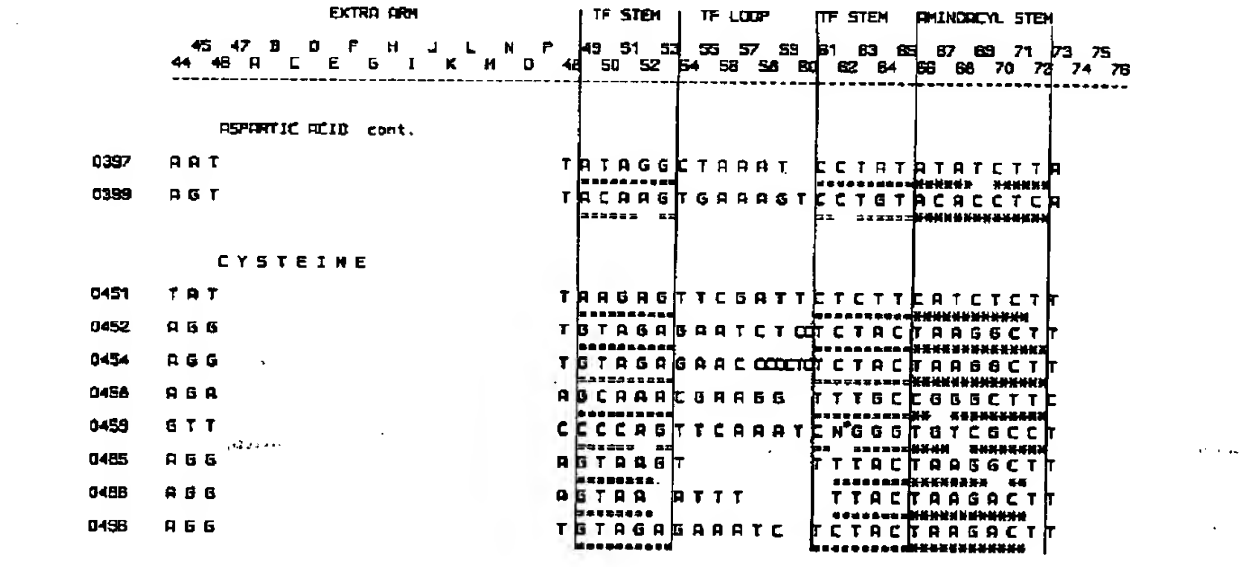
0366/0 ALIGNMENT IS ARBITRARY

0396/0 ALIGNMENT IS ARBITRARY

Nucleic Acids Research

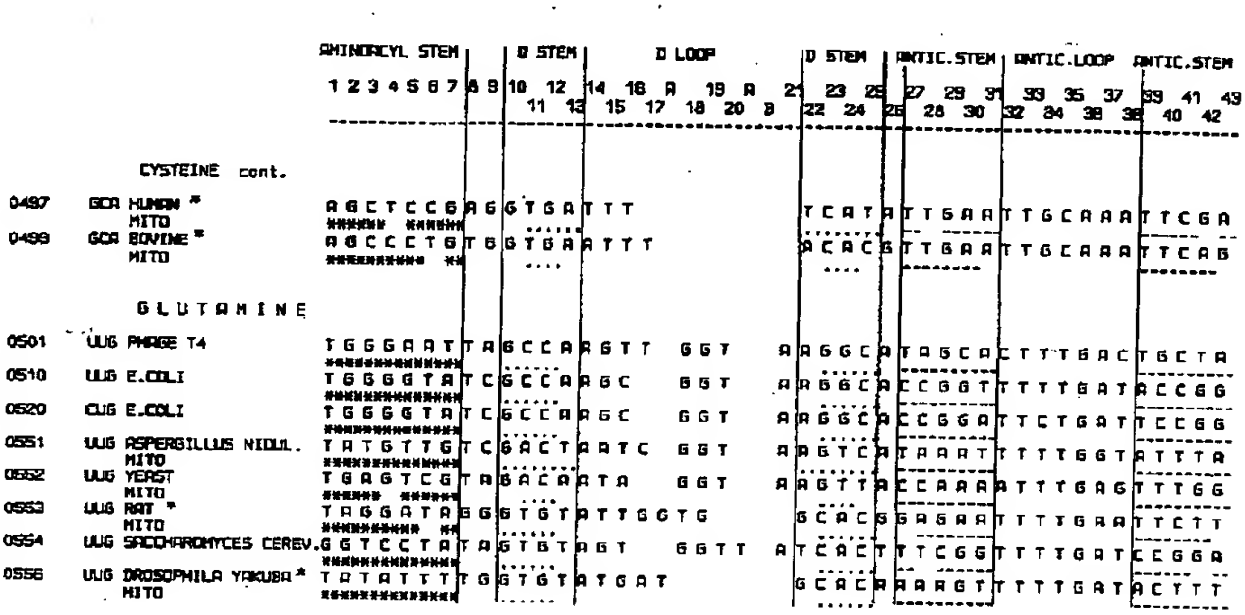


0397 S.ANDERSON ET AL. (1981) NATURE 290, 457-465
0399 S.ANDERSON, A.R.L.BECROFT,JR, A.R.COULSON, I.C.EPERON, F.SANGER, J.G.YOUNG (1982) J.MOL.BIOL. 156, 683-717
0451 J.L.BOS, H.A.DEKKUM, B.VAN DER HORST, P.BORST (1979) NUCL. ACIDS RES. 6, 3255-3266
R.E.BERLANI, C.PENTELLA, B.MACINO, A.TZAGOLOFF (1980) J. BACT. 148, 1086-1097
0452 F.CANTATORE, C.DE BENEDETTO, B.GADALETA, R.GALLERANI, A.M.KROON, H.HOLTROP, C.LANAVE, G.PEPE, G.QUAGLIARIELLO, C.SACCONE, E.SBISA (1982) NUCL. ACIDS RES. 10, 3279-3289
0454 H.TAIRA,E.YOSHIDA,S.KUWAZAWA,Y.YAMIZUMA,K.KOIKE(1983) NUCL.ACIDS RES. 11, 1635-1643
0458 J.F.M.WOOD, D.P.RA, B.R.WILSON, B.A,ROE (1983) NUCL. ACIDS RES. 11, 4977-5004
0459 X.MOLBOHM,W.HOTTORLEY,P.R.WHITFELD(1983) NUCL.ACIDS RES. 11, 8347-8354
0485 H.H.L.DE BRUIJN(1983) NATURE 304, 234-241
0486 D.O.CLARY, D.R.WOLSTENHOLME (1983) NUCL. ACIDS RES. 11, 6859-6872
0498 R.J.KIBB, R.A.VAN ETTEN, C.T.WRIGHT, N.G.WALBERG, D.A.CLAYTON (1981) CELL 26, 167-180

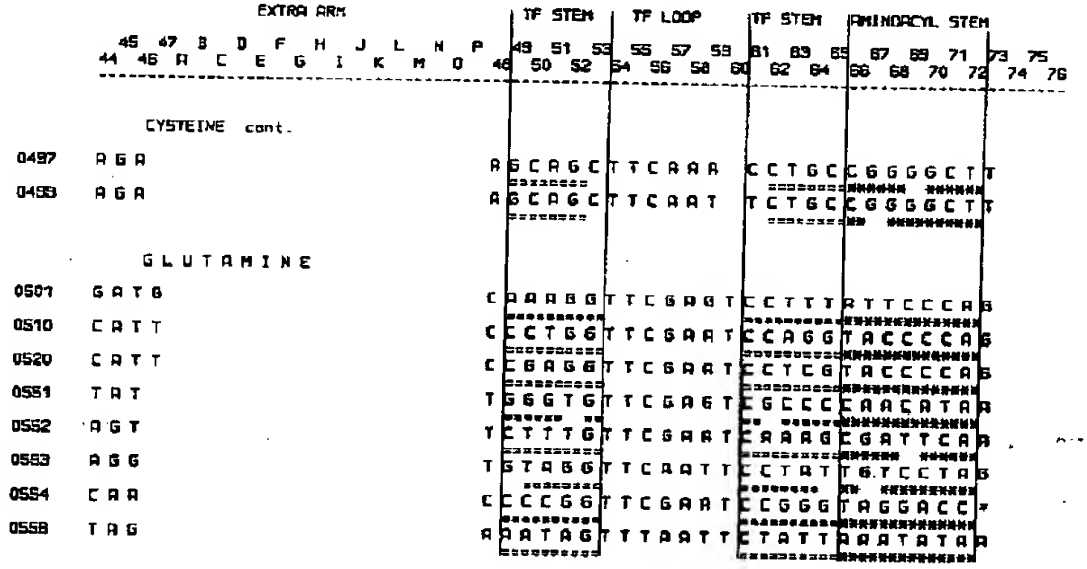


0397/0 ALIGNMENT IS ARBITRARY
0399/0 ALIGNMENT IS ARBITRARY
0452/0 ALIGNMENT IS ARBITRARY
0452/35 IS REPORTED TO BE A IN B.GROSSKOPF,H.FELDMANN(1981) CURR.GEN. 4, 191-196
0454/0 ALIGNMENT IS ARBITRARY
0458/0 ALIGNMENT IS ARBITRARY
0459/62 N IS EITHER C OR T
0485/0 ALIGNMENT IS ARBITRARY
0486/0 ALIGNMENT IS ARBITRARY
0496/0 ALIGNMENT IS ARBITRARY

Nucleic Acids Research



0497 S.ANDERSON ET AL. (1981) NATURE 290, 457-465
0499 S.ANDERSON, A.H.L.DEBRUIJN, A.R.COULSON, I.C.EPERON, F.SANGER, I.G.YOUNG (1982) J.MOL.BIOL. 156, 683-717
0501 E.FUKADA, J.ABELSON (1980) J. MOL. BIOL. 139, 377-391
0510 N.NAKAJIMA, H.OZEKI, Y.SHIMURA (1981) CELL 23, 239-249
0520 N.NAKAJIMA, H.OZEKI, Y.SHIMURA (1981) CELL 23, 239-249
0551 H.D.KOECHEL, C.R. LAZARUS, W.BASAK, B.KUENZEL (1981) CELL 23, 625-633
0552 R.E.BERLANI, G.G.BONITZ, G.CORUZZI, A.NOBREGA, A.TZAGOLOFF (1980) NUCL. ACIDS RES. 8, 5017-5030
0553 F.CANTATORE, C.DE BENEDETTO, G.GADALETA, R.GALLERANI, A.M.KROON, N.HOLTROP, C.LANAVE, G.PEPE, C.QUAGLIARIELLO, C.SACCONE, E.SBISA (1982) NUCL. ACIDS RES. 10, 3279-3287
0554 D.TSCHUDPEN,J.CARBON(1982) J.MOL.BIOL. 156,293-307
0556 D.O.CLARY, J.R.GROGAN, E.C.GARFIN, C.A.-R.FAURON, D.R.WOLSTENHOLME (1982) NUCL. ACIDS RES. 10, 6419-6437



0497/0 ALIGNMENT IS ARBITRARY
0499/0 ALIGNMENT IS ARBITRARY
0553/0 ALIGNMENT IS ARBITRARY

0554/73 THE 3'END OF THE TRNA WAS NOT DETERMINED
0554/0 ALIGNMENT IS ARBITRARY

Nucleic Acids Research

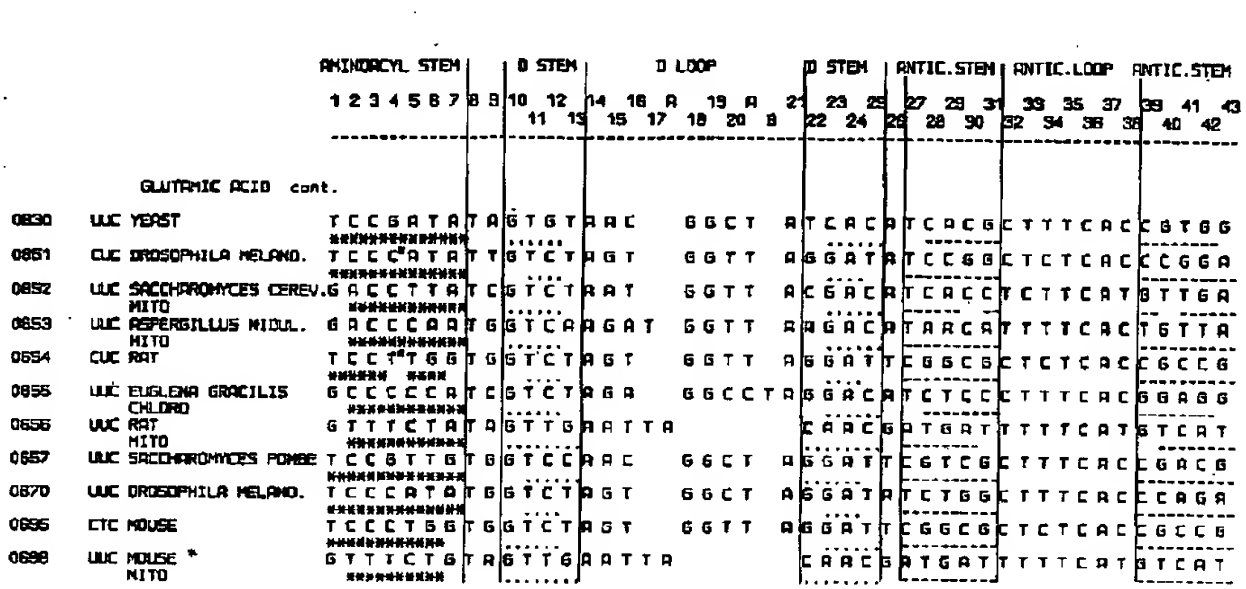
			AMINOACYL STEM								D STEM			D LOOP							D STEM			ANTIC.STEM			ANTIC.LOOP						ANTIC.STEM		
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	A	18	19	20	B	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
	GLUTAMINE cont.																																													
0557	UUG	EUGLENA GRACILIS CHLORO	T	G	A	G	G	C	G	T	A	G	C	C	A	A	G	T			G	G	T		A	A	G	G	C	A	A	C	G	G	G	T	T	T	T	G	G	C	C	C	T	G	T
0558	UUG	TOBACCO CHLORO	T	G	G	G	C	G	T	G		G	C	C	A	A	G	T			G	G	U		A	A	G	G	C	A	A	C	G	G	G	T	T	T	T	G	G	T	C	C	C	G	C
0559	UUG	NICOTIANA TABACUM CHLORO	T	G	G	G	C	G	T	G		G	C	C	A	A	G	T			G	G	T		A	A	G	G	C	A	A	C	G	G	G	T	T	T	T	G	G	T	C	C	C	G	C
0567	UUG	RATTUS NORVEGICUS MITO	T	A	G	G	A	T	A	G	G	G	T	G	T	A	T	T			G	G	T		A	G	C	A	C	G	G	A	A	A	A	T	T	T	T	G	A	A	T	T	C	T	T
0596	UUG	MOUSE MITO	T	A	G	G	A	T	A	A	G	G	T	G	T	T	T	A			G	G	T		G	C	A	C	G	G	A	G	A	A	A	T	T	T	T	G	A	A	T	T	C	T	T
0597	CUG	HUMAN * MITO	T	A	G	G	A	T	G	G	G	G	T	G	T	G	A	T	A	G	G	T	G			G	C	A	C	G	G	A	G	A	A	T	T	T	T	G	G	A	T	T	C	T	C
0598	UUG	HUMAN	G	G	T	T	C	C	A	T	G	G	T	G	T	A	A	T			G	G	T	T		A	G	C	A	C	T	C	T	G	G	A	C	T	C	T	G	A	A	T	C	C	A	G
0599	UUC	BOVINE MITO	T	A	G	A	A	T	T	T	G	G	T	G	T	A	A	T	T		G	G	G			A	G	C	A	C	G	A	A	G	A	G	T	T	T	T	G	G	A	T	T	C	T	T
	GLUTAMIC ACID																																													
0609	UUC	E.COLI	G	T	C	C	C	C	T	T	C	G	T	C	T	A	G	A			G	G	C	C	C	A	G	G	A	C	A	C	C	G	C	C	C	T	T	T	C	A	C	G	G	C	G	G
0610	UUC	BACILLUS SUBTILIS	G	G	C	C	C	G	T	T	G	G	T	C	A	A	G	C			G	G	T	T		A	A	G	A	C	A	C	C	G	C	C	C	T	T	T	C	A	C	G	G	C	G	G

0557 G.D.KARABIN, R.B.HALLICK (1981) J. BIOL. CHEM. 256, 5510-5518
0558 H.OENO,K.SUGIURA(1983) NUCLEIC ACIDS RES. 11, 6407-6414
0559 H.OENO,K.SUGIURA(1983) NUCL.ACIDS RES. 11, 6407-6414
0567 D.R.WOLSTENHOLME, C.R.-R.FARRON, J.R.GODDARD (1982) GENE 20, 63-69
0596 B.J.BIBB, R.A.VAN ETTEN, C.T.WRIGHT, M.W.WALBERG, D.A.CLAYTON (1981) CELL 26, 167-180
0597 S.ANDERSON ET AL. (1981) NATURE 290, 457-465
0598 K.I.ROY, H.COOKE, R.BUCKLAND (1982) NUCL. ACIDS RES. 10, 7313-7322
0599 S.ANDERSON, A.H.L.DEBRUIJN, A.R.COULSON, I.C.EPLEN, F.SANGER, I.G.YOUNG (1982) J. MOL. BIOL. 156, 683-717
0609 J.BROSIUS,T.J.DULL,D.D.SLEETER,H.F.NOLLER(1981) J.MOL.BIOL. 148, 107-127
E.A.MORGAN,T.IKEMURA,L.E.POST,S.NOMURA(1980) IN:TRANSFER RNA, D.SOELL, J.ABELSON,P.SCHIMMEL (ED),COLD SPRING HARB. 259-266
0610 Y.YAMADA, A.ONKI, H.ISHIKURA (1983) NUCL. ACIDS RES. 11, 3037-3045
C.J.GREEN, B.G.VOLD (1983) NUCL. ACIDS RES. 11, 5763-5774

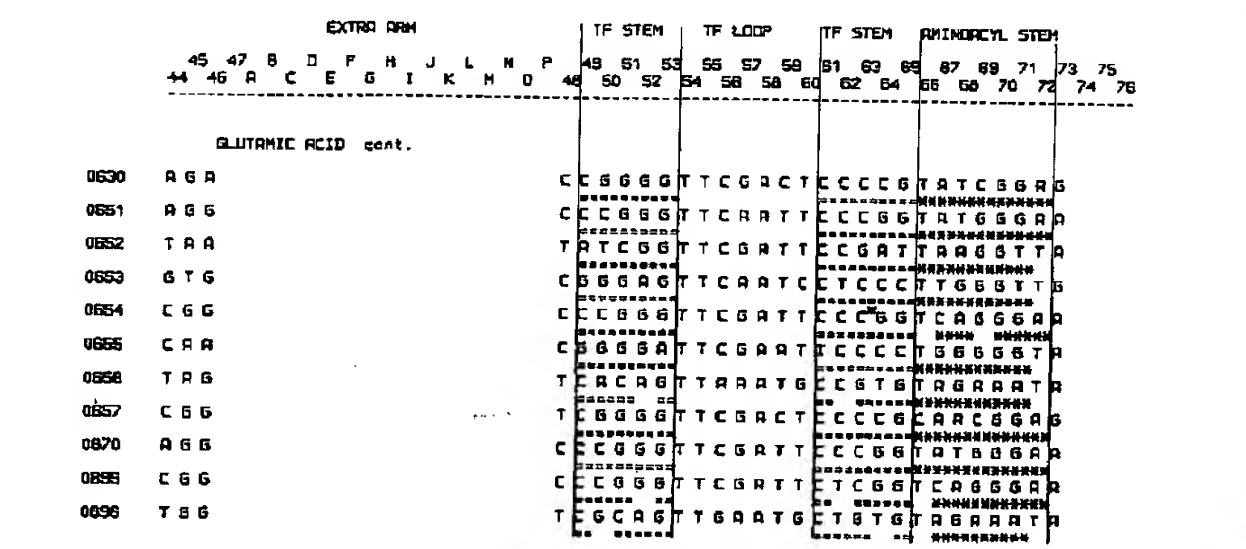
		EXTRA ARM																									TF STEM			TF LOOP			TF STEM			AMINOACYL STEM										
		44	45	46	A	47	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	48					49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76
	GLUTAMINE cont.																																															
0557		C	A	T	T																			C	G	G	A	G	G	T	T	C	G	A	A	T	C	C	T	C	C	C	G	C	C	T	C	A	G
0558		T	A	T	T																			C	G	G	A	G	G	T	T	C	G	A	A	T	C	C	T	T	C	C	G	T	C	C	C	A	G
0559		T	A	T	T																			C	G	G	A	G	G	T	T	C	G	A	A	T	C	C	T	T	C	C	G	T	C	C	C	A	G
0567		A	G	G																				T	G	T	A	G	G	T	T	C	A	A	T	T	C	C	T	A	T	T	G	T	C	C	T	A	G
0596		A	A	G																				T	G	T	A	G	G	T	T	C	A	A	T	T	C	C	T	A	T	T	G	T	C	C	T	A	G
0597		A	G	G																				G	A	T	G	G	G	T	T	C	G	A	T	T	C	T	C	A	T	A	G	T	C	C	T	A	G
0598		C	G	A																				T	C	C	G	A	G	T	T	C	A	A	A	T	C	T	C	G	G	T	G	G	A	A	C	C	T
0599		A	G	G																				A	G	T	A	G	G	T	T	C	G	A	T	T	C	C	T	A	T	A	G	T	T	C	T	A	G
	GLUTAMIC ACID																																														
0609		C	A	G	G	G	G	G	T	T	C	G	A	A	T	C	C	C	C	T	A	G	G	G	G	G	A	C	G																					
0610		C	A	C	G	G	G	T	T	C	G	A	A	T	C	C	C	G	T	A	C	G	G	G	T	C	A																					

0997/0 ALIGNMENT IS ARBITRARY

Nucleic Acids Research

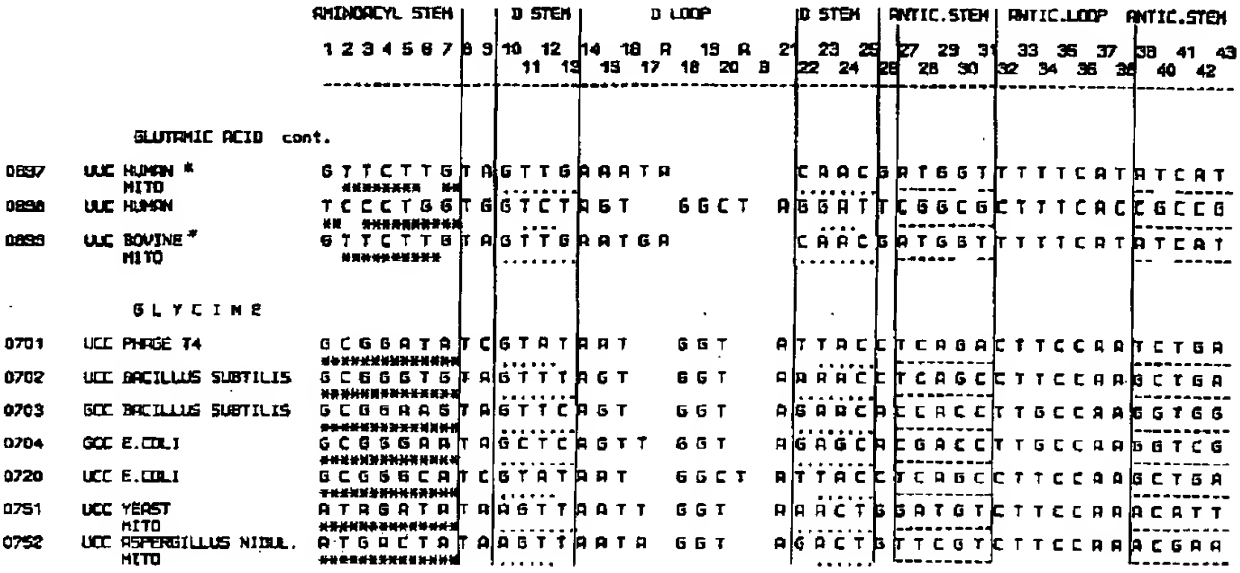


0630 A.EIGEL, J.OLAH, H.FELDMANN (1981) NUCL. ACIDS RES. 9, 2961-2970
H.FELDMANN,J.OLAH,H.FRIEDENREICH(1981) NUCL.ACIDS RES. 9, 2949-2959
A.EIGEL, H.FELDMANN (1982) EMBO-J. 1, 1245-1250
0651 M.A.MOSBACH, A.SILBERKLANG, B,J.MC CARTHY (1980) CELL 21, 169-178
0652 F.G.MOBREGA, A.TZAGOLOFF (1980) FEBS LETTERS 113, 52-54
0653 A.G.KOECHEL, C.A.LAZARUS, H.HAGAN, H.KUENTZEL (1981) CELL 23 625-633
0654 T.SEKIYA, Y.KUCHINO, S.NISHIMURA (1981) NUCL. ACIDS RES. 9, 2239-2250;
K.SHIBUYA, S.NODUCHI, S.NISHIMURA, T.SEKIYA (1982) NUCL. ACIDS RES. 10, 4441-4448
0655 A.J.HOLLINGSWORTH, R.B.WALLICK (1982) J. BIOL. CHEM. 257, 12795-12799
0656 G.GOERTZ,H.FELDMANN(1982) CURR.GEN. 5, 221-225
K.KOIKE,A.KOBAYASHI,K.YADINUMA,A.TAIRA,E.YOSHIDA,A.ISA(1982) GENE 20, 177-185
0657 V.GADULIN,J.RAO,S.APPEL,A.BURKE-SMITH,F.TARAD,D.SOELL(1983) NUCL.ACIDS RES. 11, 8537-8546
0670 Z.K.INDIK, K.D.TARTOF (1982) NUCL. ACIDS RES. 10, 4159-4172
0695 J.E.LODNEY,J.D.HARDING(1983) NUCL.ACIDS RES. 11,8761-8775
0696 A.J.BIBB, R.A.VAN EITEN, C.T.WRIGHT, H.K.WALKER, B.A.CLAYTON (1981) CELL 26, 167-180

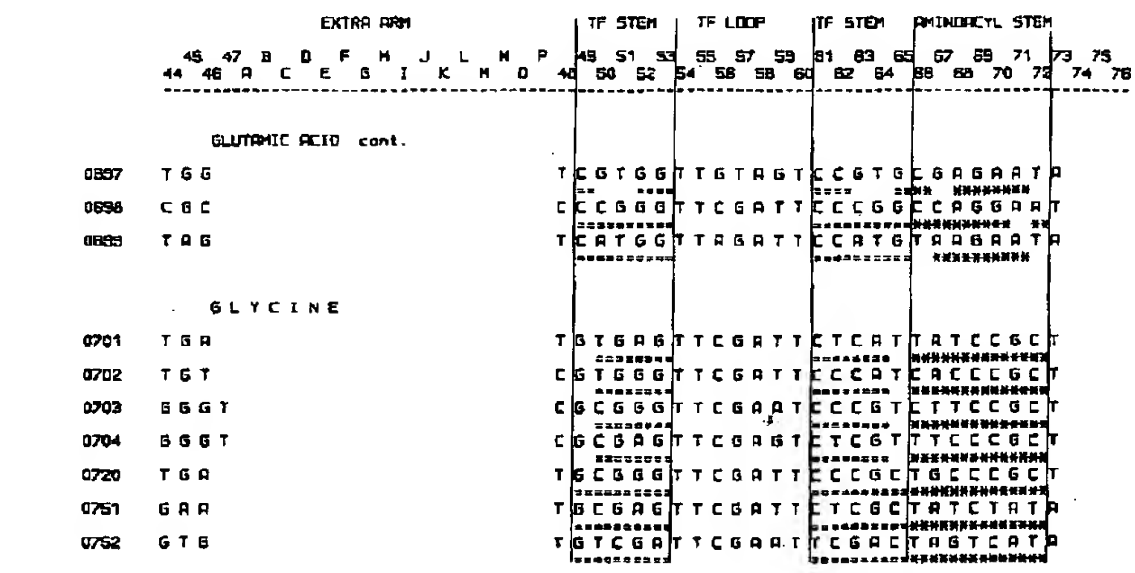


0651/4 TANDEM GENE II HAS T
0654/4 IN REPEATING UNITS C

0654/63 IN REPEATING UNITS DELETION MUTATIONS
0696/0 ALIGNMENT IS ARBITRARY



0697 S.ANDERSON ET AL. (1981) NATURE 290, 457-465
0698 J.P.BODDARD, M.SQUIRE, B.BIENZ, J.D.SMITH (1983) NUCL. ACIDS RES. 11, 2551-2562
0699 S.ANDERSON, A.H.L.DEBRUIJN, A.R.COULSON, F.C.EPERON, F.SANGER, I.G.YOUNG (1982) J. MOL. BIOL. 156, 683-717
0701 K.FUKADA, J.ABELSON (1980) J.MOL.BIOL. 139, 377-391
0702 C.J.GREEN, B.S.VOLD (1983) NUCL. ACIDS RES. 11, 5763-5774
0703 C.J.GREEN, B.S.VOLD (1983) NUCL. ACIDS RES. 11, 5763-5774
E.F.WAUROUSEK, J.M.HANSEN (1983) J. BIOL. CHEM. 258, 291-298
0704 S.D.TUCKER,A.S.BOPALAKRISHNAN,R.BOLLINGER,W.BOWMAN,E.J.MURGOLA(1982) J.BACT. 152, 771-779
0720 G.AN,J.D.FRIESEN(1980) GENE 12, 33-39
L.HUDSON, J.ROSSI, A.LANDY (1981) NATURE 294, 422-427
A.KIYASIMA,T.YOKOTA,Y.TAKEBE,H.NAKAMURA,Y.KAZIRO(1982) J.BIOCHEM. 93, 1101-1106
0751 S.B.BONITZ, A.TZAGOLOFF (1980) J. BIOL. CHEM. 255, 9075-9081
D.L.MILLER,M.C.MARTIN(1983) CURR.GEN. 4, 135-143
0752 H.G.KOECHEL, C.R.LAZARUS, W.BASAK, H.KUENTZEL (1981) CELL 23, 625-633



0697/0 ALIGNMENT IS ARBITRARY

0699/0 ALIGNMENT IS ARBITRARY

Nucleic Acids Research

			AMINOACYL STEM								D STEM			D LOOP						D STEM			ANTIC.STEN			ANTIC.LOOP				ANTIC.STEM		
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	A	18	19	A	B	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
		GLYCINE cont.																																												
0753	ACC	ASPERGILLUS NIDUL. MITO	A	C	G	G	C	T	A	T	A	A	G	T	T	A	A	T		G	G	T		A	G	A	C	T	A	C	T	T	A	G	C	T	A	C	C	A	C	C	T	A	A	G
0754	UCC	RAT	G	C	G	T	T	G	G	T	G	G	T	A	T	A	G	T		G	G	T	*G	A	G	C	A	T	A	G	C	T	G	C	C	T	T	C	C	A	A	G	C	A	G	T
0755	GCC	EUGLENA GRACILIS CHLORO	G	C	A	G	A	T	G	T	A	G	C	T	C	A	G	T		G	G	T		A	G	A	G	C	G	C	A	A	C	C	T	T	G	C	C	A	A	G	G	T	T	G
0758	UCC	EUGLENA GRACILIS CHLORO	G	C	G	G	G	T	A	T	A	G	C	T	C	A	G	T	T	G	G	T		A	G	A	G	C	G	T	G	G	T	C	C	T	T	C	C	A	A	G	T	C	C	A
0767	UCC	RAT * MITO	A	C	T	C	C	C	T	T	A	G	T	A	T	A	A	A	C	A					A	T	A	C	A	A	C	T	G	A	C	T	T	C	C	A	A	T	C	A	G	T
0785	GCC	DROSOPHILA MELANO.	G	C	A	T	C	G	G	T	G	G	T	T	C	A	G	T		G	G	T		A	G	A	A	T	G	C	T	C	G	C	C	T	G	C	C	A	C	G	C	G	G	G
0788	UCC	DROSOPHILA YAKUBA * MITO	A	T	N*	T	N*	T	A	T	A	G	T	A	T	A	A	A							A	G	T	A	T	A	T	T	T	G	A	C	T	T	C	C	A	A	T	C	A	T	A
0795	UCC	MOUSE	G	C	G	T	T	G	G	T	G	G	T	A	T	A	G	T		G	G	T	G	A	G	C	A	T	A	G	C	T	G	C	C	T	T	C	C	A	A	G	C	A	G	T
0798	UCC	MOUSE * MITO	A	C	T	C	C	C	T	T	A	G	T	A	T	A	A	T	T	A					A	T	A	T	A	A	C	T	G	A	C	T	T	C	C	A	A	T	T	A	G	T
0797	UCC	HUMAN * MITO	A	C	T	C	T	T	T	T	A	G	T	A	T	A	A	A	T	A					G	T	A	C	C	G	T	T	A	A	C	T	T	C	C	A	A	T	T	A	A	C

0753 H.G.ZOECHEL, C.A.LAZARUS, N.WASAI, H.HUENTZEL (1981) CELL 23, 625-633

0754 T.SEKIYA, Y.KUCHINO, S.NISHIMURA (1981) NUCL. ACIDS RES. 9, 2239-2250
K.OHIBUYA, S.NOGUCHI, S.NISHIMURA, T.SEKIYA (1982) NUCL. ACIDS RES. 10, 4441-4448

0755 G.D.ZARABIN, R.B.WALLICK (1983) J. BIOL. CHEM. 256, 5512-5518

0756 A.J.WILLINGSWORTH, R.B. WALLICK (1982) J. BIOL. CHEM. 257, 12795-12799

0757 P.CANTATORE, C.DE BENEDETTO, G.GADALETA, R.GALLERANI, A.A.KROON, A.HOLTROP, C.LANAVE, O.PEPE, C.QUALIGARIELLO, C.SACCONE, E.SBISA (1982) NUCL. ACIDS RES. 10, 3279-3289
R.GROSSKOPF, H.FELDMANN (1981) CURR. GEN. 4, 151-158

0785 N.D.HERSHEY, M.DAVIDSON (1980) NUCL. ACIDS RES. 8, 4899-4910

0788 D.O.CLARY, J.A. WAHLEITHNER, D.R.WOLSTENHOLME (1983) NUCL. ACIDS RES. 11, 2411-2423

0795 J.C.HU, B.O.COTE, E.LUND, J.E.DAHLBERG (1983) NUCL. ACIDS RES. 11, 4309-4321
J.E.LOONEY,J.D.HARDIN8(1983) NUCL.ACIDS RES. 11, 4761-4775

0796 A.J.BIBB, R.A.VAN ETTEN, C.T.WRIGHT, M.W.WALBERG, D.A.CLAYTON (1981) CELL 26, 167-180

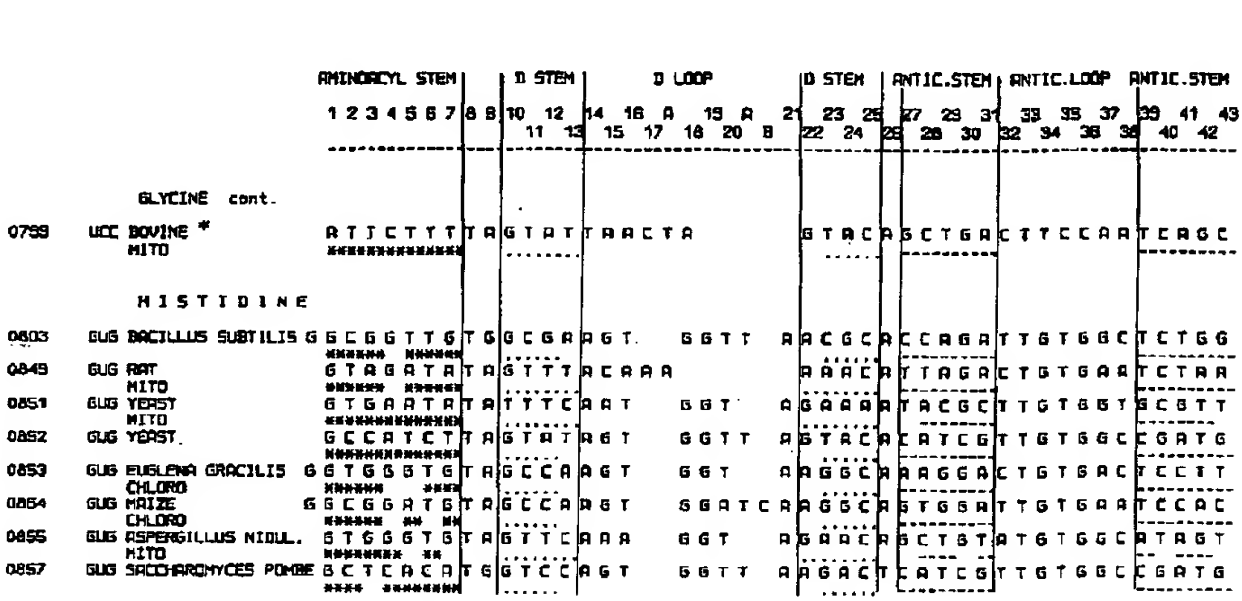
0797 S.ANDERSON ET AL. (1981) NATURE 290, 457-465

		EXTRA ARM											TF STEM			TF LOOP			TF STEM			AMINOACYL STEM						
		45	47	B	D	F	H	J	L	N	P	49	51	53	55	57	59	61	63	65	67	69	71	73	75
		44	46	A	C	E	G	I	K	M	O	48	50	52	54	56	58	60	62	64	66	68	70	72	74	76
		GLYCINE cont.																										
0753		A	T	G								T	G	T	C	G	A	T	T	C	G	A	A	T	T	C	G	A	C	T	A	G	C	C	G	T	A
0754		T	G	A								C	C	C	G	G	G	T	T	C	G	A	T	T	C	C	C	*G	G	C	C	A	A	C	G	C	*A
0755		A	T	G	C							C	A	T	G	G	G	T	T	C	G	A	G	T	C	C	C	A	T	T	A	T	C	T	G	C	T
0758		A	T	G	T							T	G	C	G	T	G	T	T	C	G	A	A	T	C	A	C	G	T	T	A	C	C	C	G	C	T
0757		T	A	A								T	T	C	T	G	A	A	A	A	A	C			T	C	A	G	A	A	G	A	G	T	A
0785		C	G	G								C	C	C	G	G	G	T	T	C	G	A	T	T	C	C	C	G	G	C	C	G	A	T	G	C	A
0786		A	G	G								T	C	T	A	T	T	A	A	T	A				A	A	T	A	G	T	A	T	A	G	A	T	A
0795		T	G	A								C	C	C	G	G	G	T	T	C	G	A	T	T	C	C	C	G	G	C	C	A	A	C	G	C	A
0796		A	G	A								T	T	C	T	G	A	A	T	A	A	A	C		C	C	A	G	A	A	G	A	G	T	A		
0797		T	A	G								T	T	T	T	G	A	C	A	A	C	A	T		T	C	A	A	A	A	A	G	A	G	T	A		

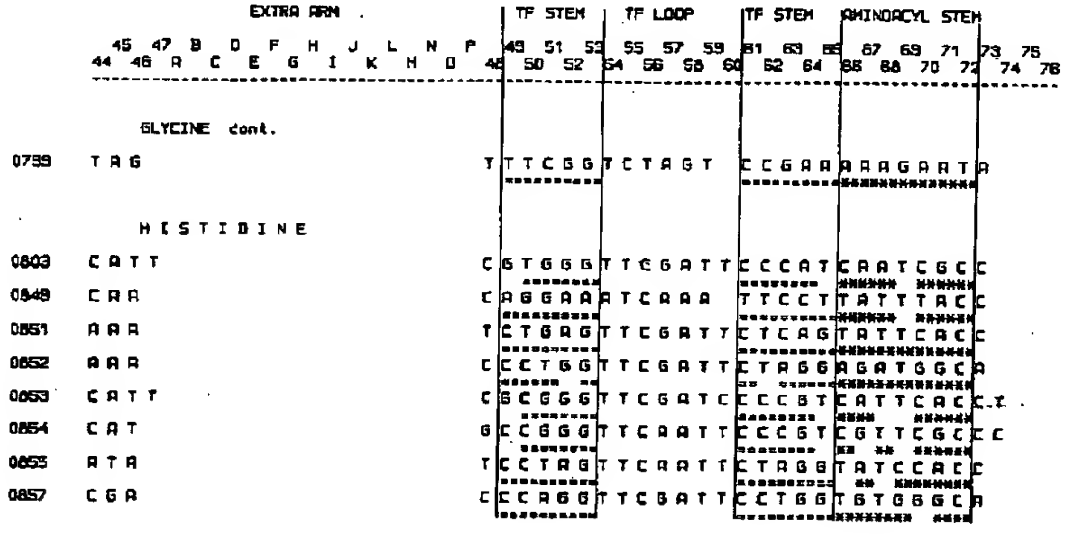
0754/20 A TO 26 IN REPEATING UNITS DELETION MUTANTS
0754/62 IN REPEATING UNIT T
0754/72 IN REPEATING UNITS T
0757/0 ALIGNMENT IS ARBITRARY
0758/0 ALIGNMENT IS ARBITRARY

0786/3 IS EITHER C OR A
0786/5 IS EITHER A OR C
0796/0 ALIGNMENT IS ARBITRARY
0797/0 ALIGNMENT IS ARBITRARY

Nucleic Acids Research

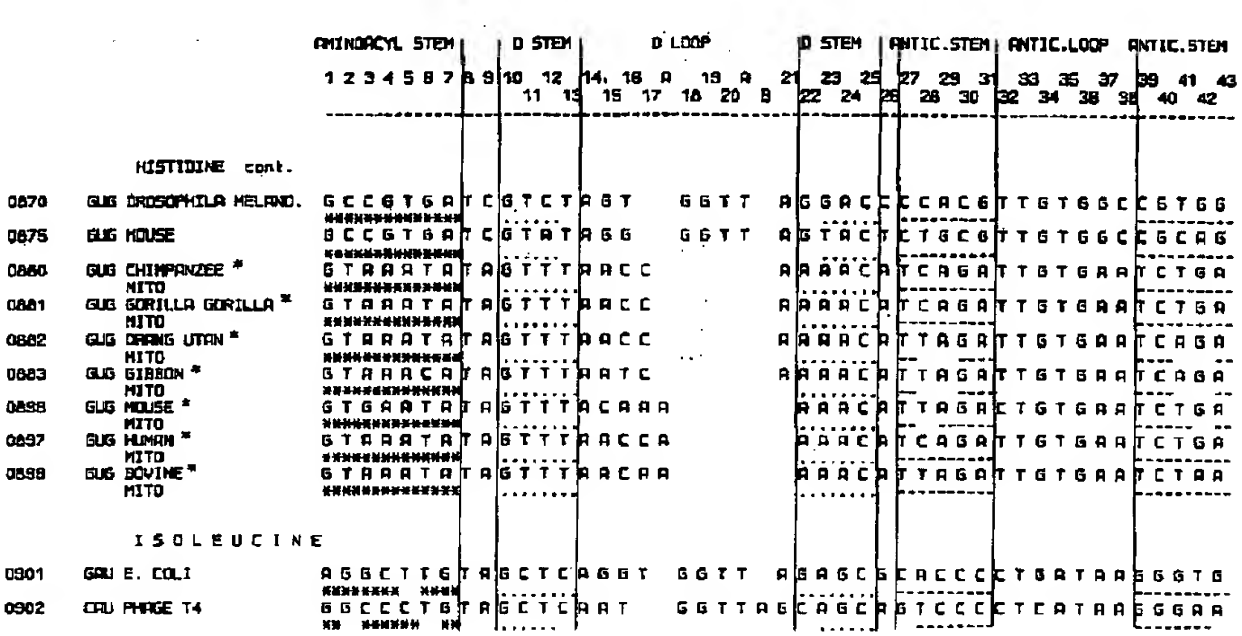


0799 S.ANDERSON, A.H.L.DEBRUIJN, A.R.COULSON, I.C.EPERON, F.SANGER, I.G.YOUNG (1982) J. MOL. BIOL. 156, 683-717
0803 C.J.GREEN, B.S.VOLD (1983) NUCL. ACIDS RES. 11, 5763-5774
0849 H.BROSKOFF, H.FELDMANN (1981) CURR. GEN. 4, 191-196
0851 J.L.BOS, K.A.OSINGA, B.VAN DER HORST, P.BORST (1979) NUCL. ACIDS RES. 6, 3255-3266;
R.E.HERLANI, C.PENTELLA, B.MACINO, A.TZAGOLOFF (1980) J. BACT. 141, 1086-1097
0852 F.DEL REY, T.F.DONAHUE, G.R.FINK (1983) J. BIOL. CHEM. 258, 8175-8182
0853 A.J.HOLLINGSWORTH, R.B. HALLICK (1982) J. BIOL. CHEM. 257, 12795-12799
0854 Z.SCHWARZ, B.O.JOLLY, A.R.STEINKOETZ, L.BOGORAD (1981) PROC. NATL. ACAD. SCI. USA 78, 3423-3427
0855 R.METZENBERG, H.G.KOECHEL, M.BAZAK, M.LUENTZEL (1982) NUCL. ACIDS RES. 10, 4783-4794
0857 V.GARDLIN, J.MAO, B.APPEL, M.BUKMER-SMITH, F.YANAO, D.SOELL (1983) NUCL.ACIDS RES. 11, 8537-8548

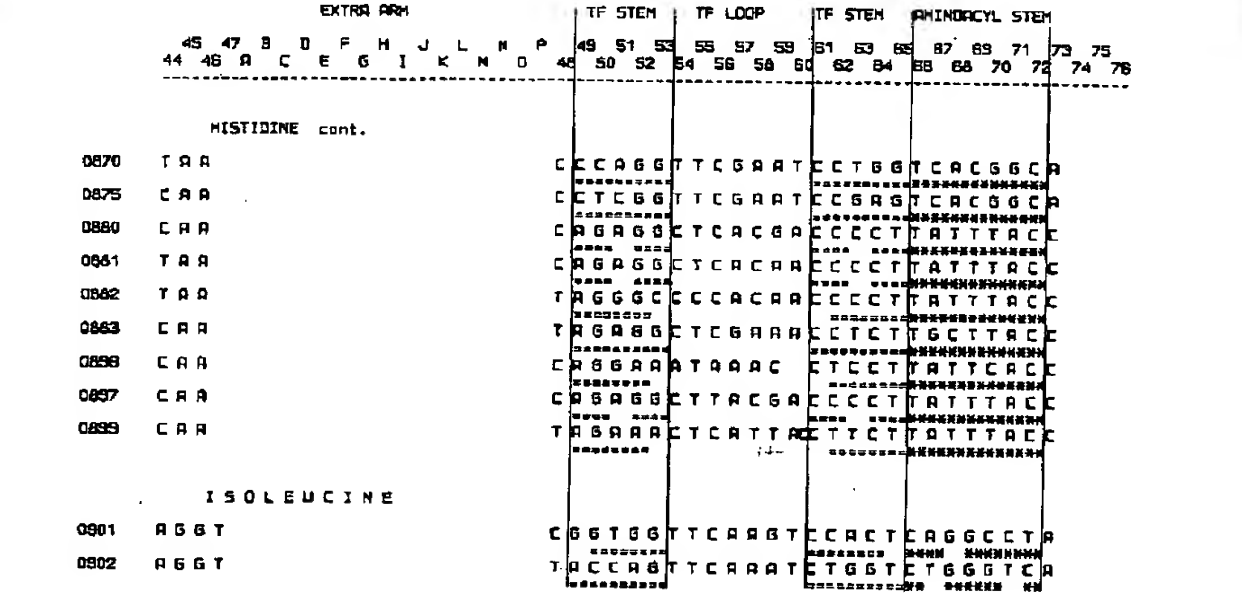


0799/0 ALIGNMENT IS ARBITRARY

Nucleic Acids Research



0870 L.COOLEY, R.APPEL, D.SOELL (1982) PROC. NATL. ACAD. SCI. USA 79, 6475-6479
0875 J.H.HAN, J.D.HARDING (1982) NUCL. ACIDS RES. 10, 4891-4900
0880 W.M.BROWN,E.M.PRAGER,A.WANG,A.C.WILSON(1982) J.MOL.EVOL. 18, 225-239
0881 W.M.BROWN,E.M.PRAGER,A.WANG,A.C.WILSON(1982) J.MOL.EVOL. 18, 225-239
0882 W.M.BROWN,E.M.PRAGER,A.WANG,A.C.WILSON(1982) J.MOL.EVOL. 18, 225-239
0883 W.M.BROWN,E.M.PRAGER,A.WANG,A.C.WILSON(1982) J.MOL.EVOL. 18, 225-2347
0898 R.J.ARGUS, R.A.VAN ETTEN, C.T.WRIGHT, M.W.VALSERO, D.A.CLAYTON (1981) CELL 24, 367-180
0897 S.ANDERSON ET AL. (1981) NATURE 290, 457-465
W.M.BROWN,E.M.PRAGER,A.WANG,A.C.WILSON(1982) J.MOL.EVOL. 18, 225-239
0899 S.ANDERSON, M.H.L.DE BRUIJN, A.R.COULSON, I.C.EPERON, F.SANGER, I.G.YOUNG (1982) J. MOL. BIOL. 156, 683-717
0901 T.SENITA, S.NISHIMURA (1979) NUCL. ACIDS RES. 6, 575-592
0902 E.FUKADA, J.ABELSON (1980) J.MOL.BIOL. 139, 377-391
G.P.RIZZARA, D.PLEWNETT, W.M.MCCLAIN (1981) PROC. NATL. ACAD. SCI 78, 889-892



0880/0 ALIGNMENT IS ARBITRARY
0881/0 ALIGNMENT IS ARBITRARY
0883/0 ALIGNMENT IS ARBITRARY
0883/0 ALIGNMENT IS ARBITRARY

0896/0 ALIGNMENT IS ARBITRARY
0897/0 ALIGNMENT IS ARBITRARY
0899/0 ALIGNMENT IS ARBITRARY

Nucleic Acids Research

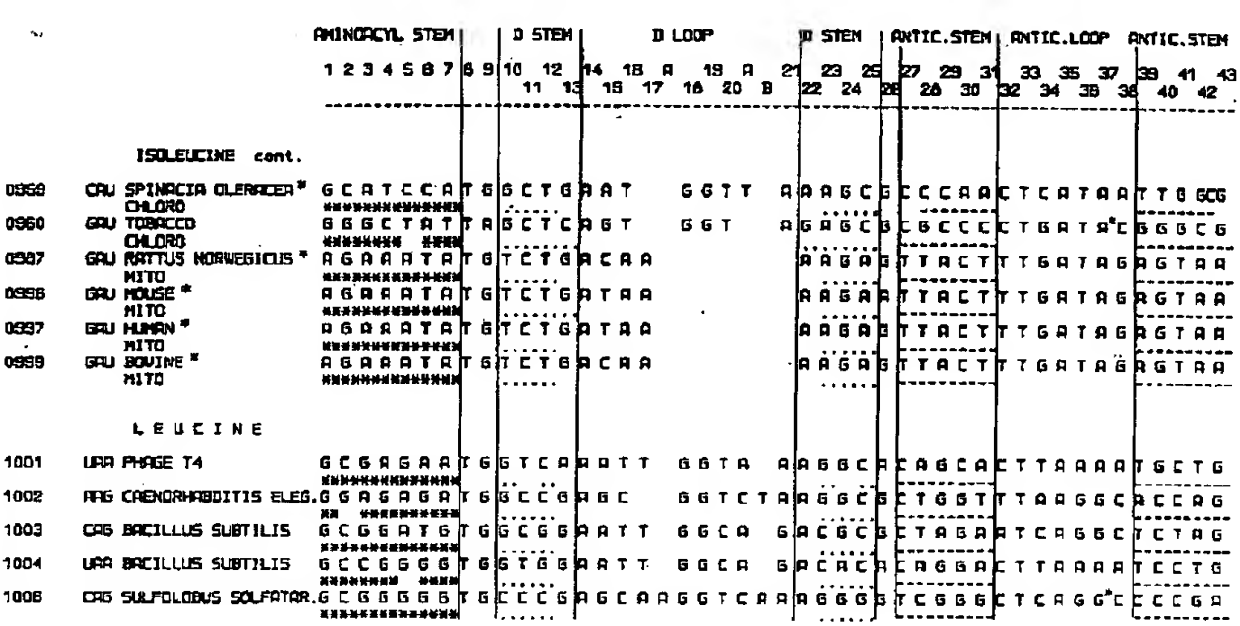
		AMINOACYL STEM	D STEM	D LOOP	D STEM	ANTIC.STEM	ANTIC.LOOP	ANTIC.STEM
		1 2 3 4 5 6 7 8 9	10 12 11 13	14 16 15 17 18 19 A 20 B 21	22 24 23 25	26 28 27 29 30 31	32 34 33 35 36 37 38	39 41 40 42 43
	ISOLEUCINE cont.							
0903	GRU BACILLUS SUBTILIS	G G G C C T G T A	G C T C A G C T	G G T T	A G A G C	G C A C G C	C T G A T A A	G C G T G
0904	CAU BACILLUS SUBTILIS	G G A C C T T T A	G C T C A G T T	G G T T	A G A G C	A G A C G G	C T C A T A A	C C G T C
0905	GAU BACILLUS SUBTILIS	G G G C C T G T A	G C T C A G C T	G G T T	A G A G C	G C A C G C	C T G A T A A	G C G T G
0950	UAU YEAST	G C T C G T G T A	G C T C A G T	G G T T	A G A G C	T T C G T G	C T T A T A*	C G C G A
0951	GAU EUGLENA GRACILIS CHLORO	G G G C T A T T A	G C T C A G T T	G G T T	A G A G C	A T A C C C	T T G A T A A	G G G T A
0952	AAU DROSOPHILA MELANO.	G G C C C A T T A	G C T C A G T T	G G T T	A G A G C	G T C G T G	C T A A T A A	C G C G A
0953	GAU ASPERGILLUS NIDUL. MITO	G G T T C T T T A	A C T T A A C C	G G T	A A A G T	G T G T T C	T T G A T A A	G G A T A
0954	GAU MAIZE CHLORO	G G G C T A T T A	G C T C A G T	G G T	A G A G C	G C G C C C	T G A T A A*	G G G C G
0955	GAU RAT " MITO	A G A A A T A T	G T C T G A C A A		A A G A	G T T A C T	T T G A T A G	A G T A A
0958	GAU DROSOPHILA YAKUBA MITO	A A T G A A T T	G C C T G A T A A		A A A G G	G T T A C C	T T G A T A G	G G T A A
0957	GAU ANACYSTIS NIDULANS	G G G C T A T T A	G C T C A G G T	G G T T	A G A G C	G C A C C C	T G A T A A	G G G T G

0903 K.LOUBINET, E.LUND, J.E.DAHLBERG (1982) NUCL. ACIDS RES. 10, 1607-1624
N.OBASANARA, S.HORIYA, H.YOSHIKAWA (1982) NUCL.ACIDS RES. 11, 6301-6318
0904 C.J.GREEN, B.S.VOLD (1983) NUCL. ACIDS RES. 11, 3763-3774
0905 C.J.GREEN, B.S.VOLD (1983) NUCL.ACIDS RES. 11, 3763-3774
0950 C.L.PEEBLES, P.GEGENHEIMER, J.ABELSON (1983) CELL 32, 525-536
0951 L.GRAF, H.KOESSEL, E.STUTZ (1980) NATURE 284, 906-910;
E.M.DROZCO, JR., N.E.RUSHLOW, J.R.DODD, R.B. MALLICH (1980) J. BIOL. CHEM. 255, 10997-11003
T.HIYATA, R.KIKUNO, Y.OSHIMA (1982) NUCL.ACIDS RES. 10, 1771-1780
0952 B.NOVERARM, S.BHARP, N.YAMADA, D.SOELL (1980) CELL 19, 889-895
A.R.ROBINSON, N.DAVIDSON (1981) CELL 23, 251-259
0953 H.B.KOECHEL, C.A.LAZARUS, H.KABAK, H.KUNTZEL (1981) CELL 27, 629-633
0954 H.KOCH, E.EDWARDS, H.KOESSEL (1981) CELL 25, 203-212;
0955 P.CANTATORE, C.DE BENEDETTO, A.GADALETA, E.GALLERANI, A.A.KADON, B.HELTROP, C.LANAVE, G.PEPE, C.QUAGLIARIELLO, C.SACCONE, S.SBISA (1982) NUCL. ACIDS RES. 10, 1279-3289
0958 D.O.CLARY, J.R.GODDARD, S.C.MARTIN, C.R.-E. FAURON, D.R.WOLSTENHOLME (1982) NUCL. ACIDS RES. 10, 6419-6437
0957 S.E.WILLIAMSON, W.F.DOOLITTLE (1983) NUCL. ACIDS RES. 11, 225-235

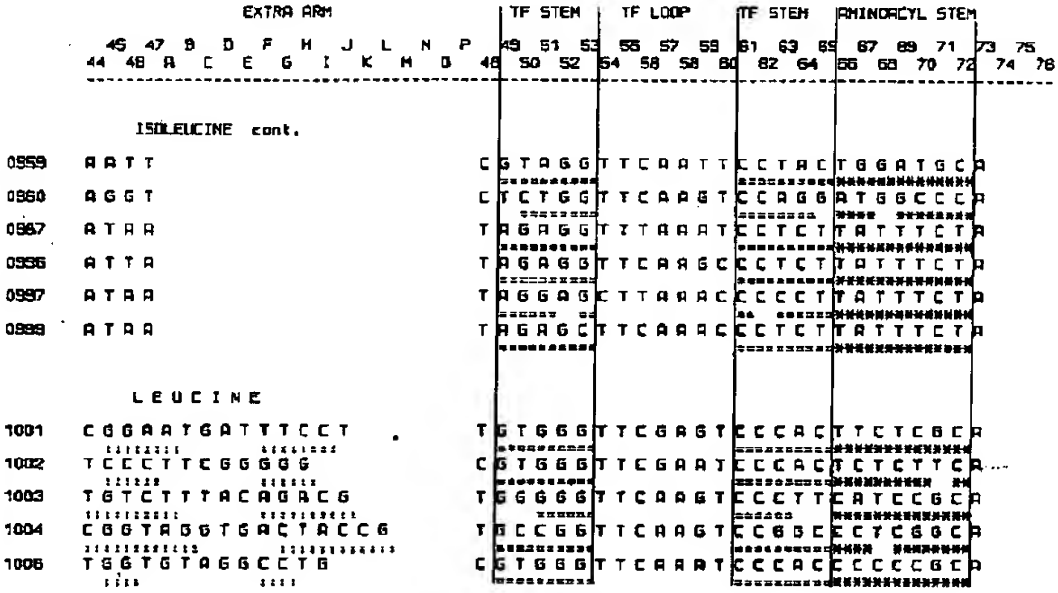
		EXTRA ARM	TF STEM	TF LOOP	TF STEM	AMINOACYL STEM
		45 47 B D F H J L N P 44 48 A C E G I K M O 46	49 51 53 50 52	55 57 59 54 56 58 60	61 63 65 62 64	67 69 71 73 75 68 66 70 72 74 76
	ISOLEUCINE cont.					
0903	A G G T	C	G G T G G	T T C G A G T	C C A C T	C A G G C C C A
0904	C G G T	C	G T A G G	T T C G A G T	C C T A C	A A G G T C C A
0905	A G G T	C	G A T G G	T T C G A G T	C C A T T	C A G G C C C A
0950	C C G T	C	G T G G G	T T C A A T C	C C C A C	C T C G A G C A
0951	A G G T	C	G C T A G	T T C A A G T	C T A G C	A T G G C C C T
0952	A G G T	C	G C G G G	T T C G A T C	C C C T C	A T G G G C C A
0953	T G T	T	C A G T G	T T C G A G T	C A C T G	A A G A A T C A
0954	A G G T	C	T C T G G	T T C A A G T	C C A G G	A T G G C C C A
0955	A T A A	T	A G A G G	T T T A A A T	C C T C T	T A T T T C T A
0956	A T T A	T	G C A G T	T T	T C T G C	A T T C A T T G
0957	A G G T	C	C C T G G	T T C A A G T	C C A G G	A T G G C C C A

0950/37 AFTER RESIDUE 37 INTERVENING SEQUENCE
0954/38 AFTER RESIDUE 34 INTERVENING SEQUENCE

P.GUILLERAUT, J.H.WEIL (1982) NUCL. ACIDS RES. 10, 1653-1659
0955/0 ALIGNMENT IS ARBITRARY

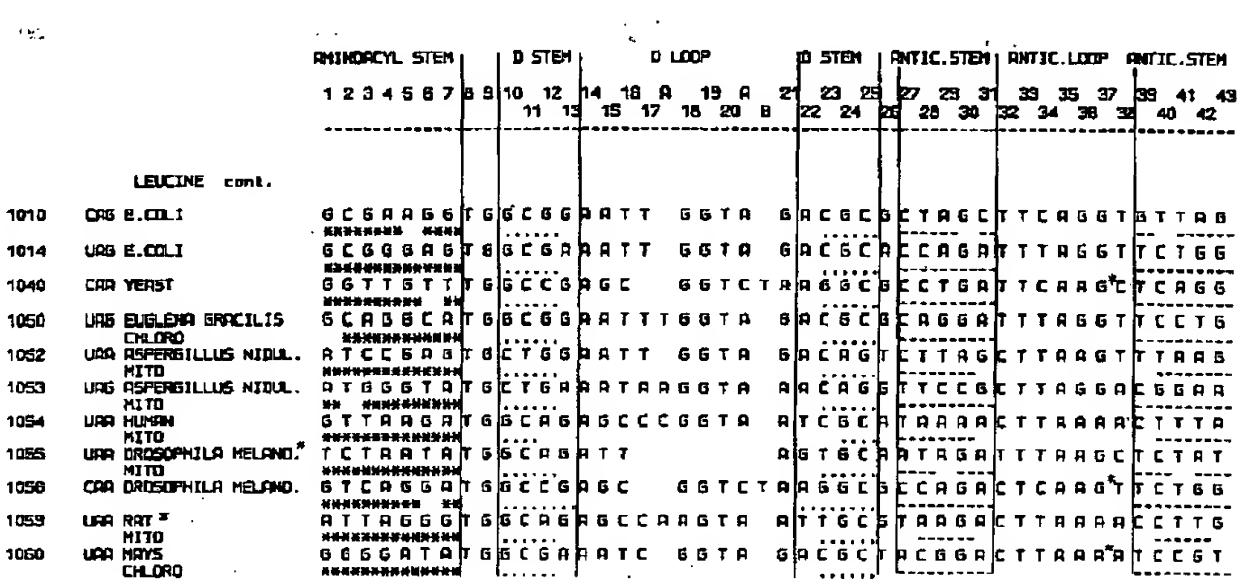


0959 B.A.KAMKAR,B.G.DUBOCK(1982) J.BIOL.CHEM. 257, 11191-11194
B.G.DRUISSEN,B.M.GREENBERG,G.ZURAWSKI,D.A.PRESCOTT,R.B.HALLICK(1983) CELL 35, 315-323
0960 F.TAKAIWA, M.SUGIURA (1982) NUCL. ACIDS RES. 10, 2665-2676
0987 D.R.WOLSTENHOLME, C.H.-M.FAURON, J.M.GOUDARD (1982) GENE 20, 43-69
0996 A.J.BIBB, R.A.VAN ETTEN, C.T.WRIGHT, M.W.WALBERG, D.A.CLAYTON (1981) CELL 26, 167-186
0997 S.ANDERSON ET AL. (1981) NATURE 290, 457-465
0999 S.ANDERSON, M.H.L.DEBRUIJN, A.R.COULSON, I.C.EPERON, F.SANGER.
I.B.YOUNG (1982) J. MOL. BIOL. 156, 683-717
1001 K.FUKADA, J.ABELSON (1980) J.MOL.BIOL. 139, 377-391
1002 T.A.TRANQUILLA, R.CORTESE, D.HELTON, J.D.SMITH (1982) NUCL. ACIDS RES. 10, 7919-7934
1003 C.J.GREEN, B.S.GOLD (1983) NUCL. ACIDS RES. 11, 5761-5774
1004 C.J.GREEN, B.S.GOLD (1983) NUCL. ACIDS RES. 11, 5761-5774
1006 S.P.KAINE, R.GUPTA, C.R.WOESE (1983) PROC. NATL. ACAD SCI. USA 80, 3309-3312

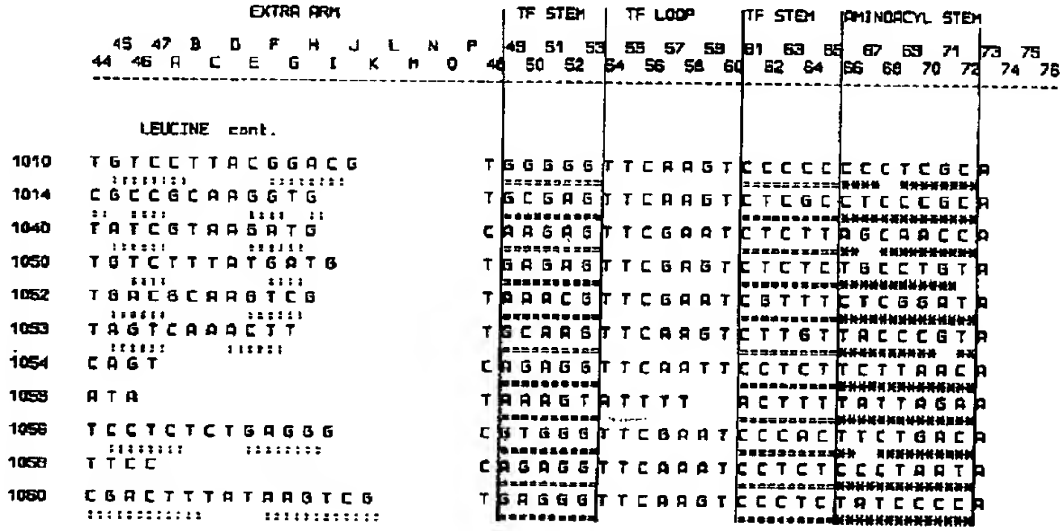


0959/0 ALIGNMENT IS ARBITRARY
0960/32 INTERVENING SEQUENCE AFTER THIS POSITION
0987/0 ALIGNMENT IS ARBITRARY
0996/0 ALIGNMENT IS ARBITRARY
0997/0 ALIGNMENT IS ARBITRARY
0999/0 ALIGNMENT IS ARBITRARY
1006/37 POSSIBLE SITE OF INTRON CLEAVAGE

Nucleic Acids Research



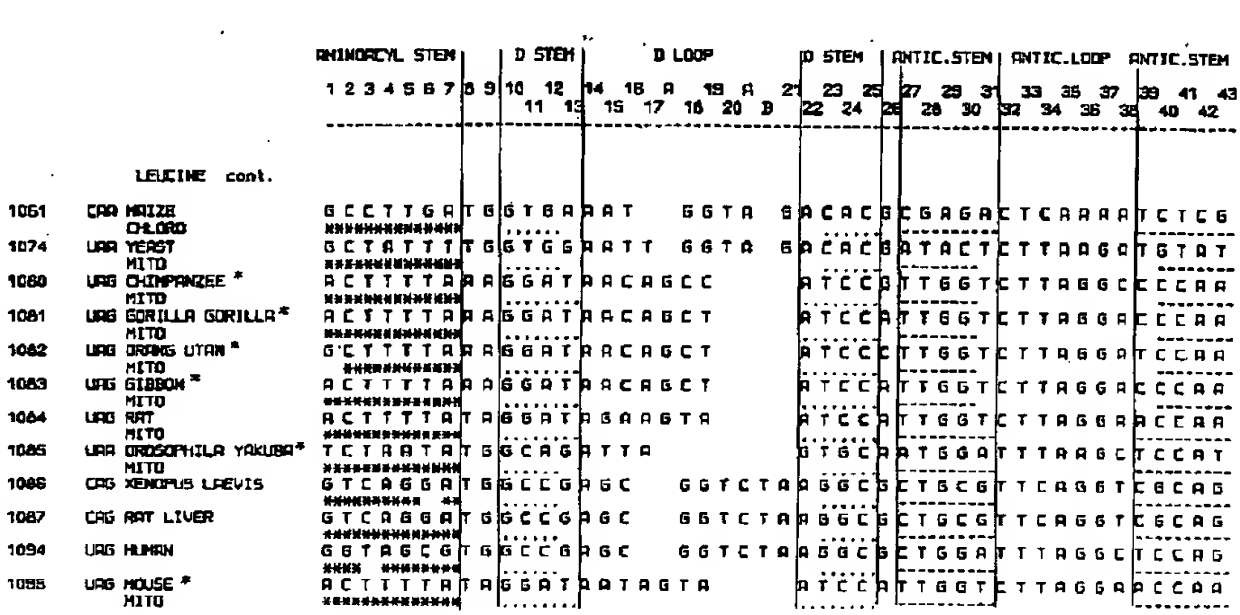
1010 G.QUESTER, R.K.CAMPEN, J.G.HOLMES (1981) NUCL. ACIDS RES. 9, 2121-2139
1014 M.NARAJIMA, M.OZEKI, Y.SHIMURA (1981) CELL 23, 239-249
1040 D.M.STANDRING, A.VENEDAS. U.J.KUTTEX (1981) PROC. NATL. ACAD. SCI. 78, 3743-3747
A.J.NEWMANK,R.C.OBDEN,J.ABELSON(1983) CELL 35, 117-123
1050 S.M.OBOZICO, JR., K.B.HALLICK (1982) J. BIOL. CHEM. 257, 3265-3275
M.GROISSER,B.B.GREENBERG,G.ZURAWSKI,D.D.PRESCOTT,R.B.HALLICK(1983) CELL 35, 815-828
1052 H.G.KOECHEL, C.M.LAZARUS, M.BABAK, M.KUENTZEL (1981) CELL 23, 625-633
1053 H.G.KOECHEL, C.M.LAZARUS, M.BABAK, M.KUENTZEL (1981) CELL 23, 625-633
1054 I.C.KFERON, S.ANDERSON, B.P.BIERLICH (1980) NATURE 286, 460-467
ANDERSON ET AL. (1981) NATURE 290, 457-465
1055 M.M.L.DE BRUIJN(1983) NATURE 304, 234-241
1056 M.M.ROBINSON, M.DAVIDSON (1981) CELL 23, 251-259
1059 G.SACCONE, P.CANTATORE, G.GADALETA, M.GALLERANI. C.LANAVE, G.PEPE A.M.KROON (1982) NUCL. ACIDS RES. 9, 4139-4148
1060 M.STEINMETZ, T.J.GUBBINS, L.BOGORAD (1982) NUCL. ACIDS RES. 10, 3027-3037



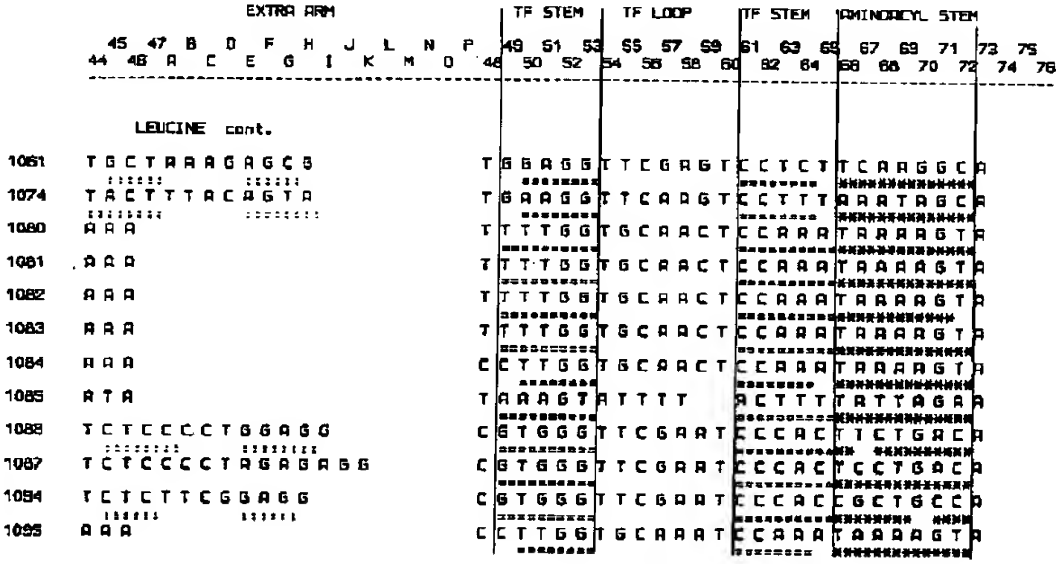
1040/37 AFTER RESIDUE 37 INTERVENING SEQUENCE
1055/0 ALIGNMENT IS ARBITRARY
1056/37 AFTER RESIDUE 37 INTERVENING SEQUENCE

1059/0 ALIGNMENT IS ARBITRARY
1060/37 AFTER RESIDUE 34 INTERVENING SEQUENCE

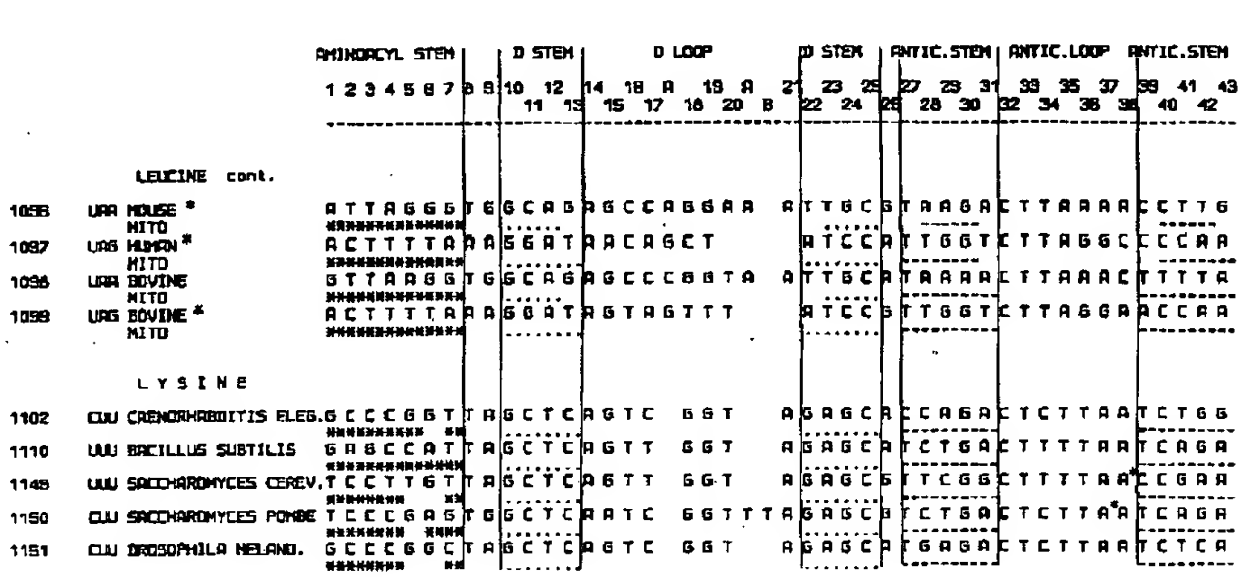
Nucleic Acids Research



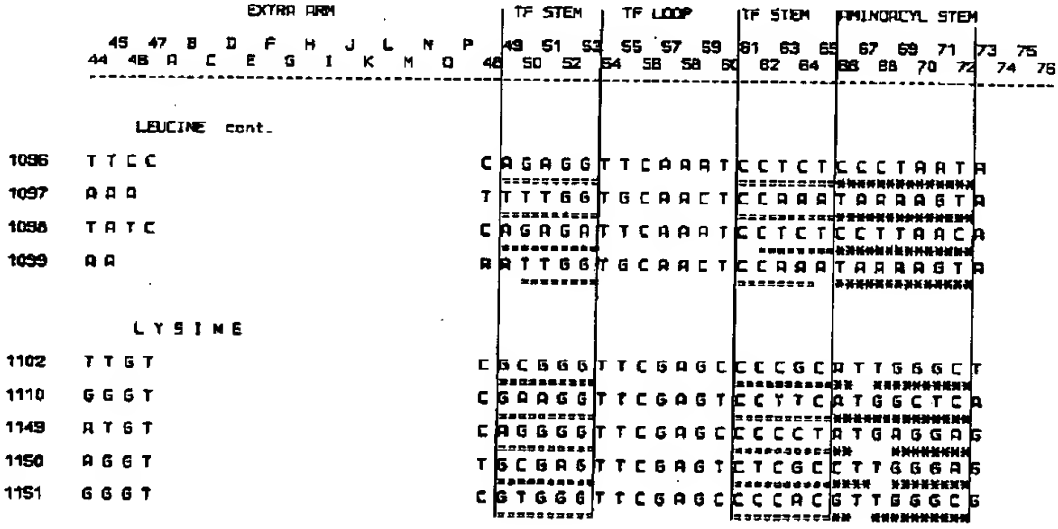
1061 A.A.STEINMETZ, E.T.KRENBERG, Z.SCHWARZ, E.J.GUBBINS, L.BOROWAD (1983) J.BIOL.CHEM. 258, 5503-5511
1074 B.BEALAMI, S.G.SONITZ, G.COROZZI, A.MORREGA, A.TZAGOLOFF (1980) NUCL. ACIDS RES. 8, 5017-5030
1080 W.M.BROWN,E.M.PRAGER,A.WANG,A.C.WILSON(1982) J.MOL.EVOL. 18, 225-239
1081 W.M.BROWN,E.M.PRAGER,A.WANG,A.C.WILSON(1982) J.MOL.EVOL. 18, 225-239
1082 W.M.BROWN,E.M.PRAGER,A.WANG,A.C.WILSON(1982) J.MOL.EVOL. 18, 225-239
1083 W.M.BROWN,E.M.PRAGER,A.WANG,A.C.WILSON(1982) J.MOL.EVOL. 18, 225-239
1084 R.GROSSKOPF, H.FELDMANN (1981) CURR. GEN. 4, 191-196
1085 D.O.CLARY, D.R.WOLSTENHOLME (1983) NUCL. ACIDS RES. 11, 4211-4227
1086 G.GALLI, H.HOFSTETTER, M.L.BIRNSTIEL (1981) NATURE 294, 626-631
1087 D.R.MACDONEL,R.A.HAAS,K.P.DOLAN,D.GRUNBERGER(1983) NUCL.ACIDS RES. 11, 3609-3624
1094 K.L.ROY, M.CROKE, R.BUCKLAND (1982) NUCL. ACIDS RES. 10, 7313-7322
1095 M.J.BIBB, R.A.VAN ETTEN, C.T.WRIGHT, M.W.WALBERG, D.A.CLAYTON (1981) CELL 26, 167-180



1080/0 ALIGNMENT IS ARBITRARY
1081/0 ALIGNMENT IS ARBITRARY
1082/0 ALIGNMENT IS ARBITRARY
1083/0 ALIGNMENT IS ARBITRARY
1085/0 ALIGNMENT IS ARBITRARY
1095/0 ALIGNMENT IS ARBITRARY



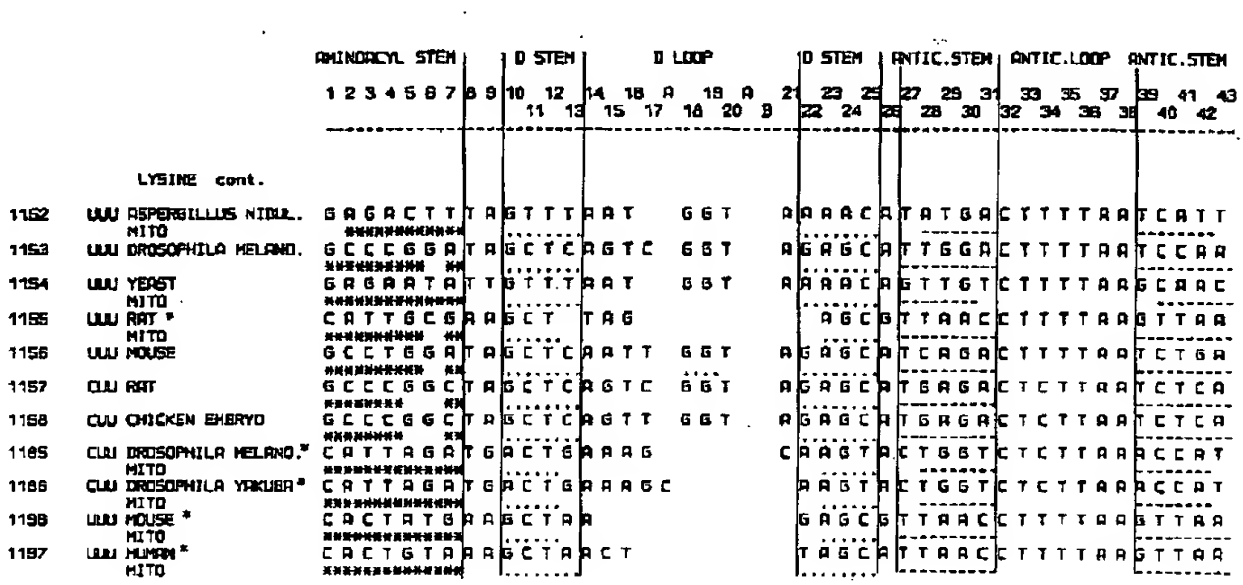
1096 B.G.BARR,A.A.VAN ETTEN,C.T.WRIGHT,A.M.WALBERG,D.A.CLAYTON(1983) CELL 26, 167-180
1097 S.ANDERSON ET AL. (1981) NATURE 290, 457-465
W.A.BROWN,E.A.PRAGER,A.WANG,A.C.WILSON(1982) J.MOL.EVOL. 18, 225-239
1098 S.ANDERSON, A.N.L.DEBRUIJN, A.R.COULSON, I.C.EPERON, F.SANGER, I.G.YOUNG, J. MOL. BIOL. 156, 683-717
1099 S.ANDERSON, A.N.L.DEBRUIJN, A.R.COULSON, I.C.EPERON, F.SANGER, I.G. YOUNG (1982) J. MOL. BIOL. 156, 683-717
1102 T.A.TRANQUILLA, R.CORTESE, D.MELTON, J.D.SMITH (1982) NUCL. ACIDS RES. 10, 7919-7934
1110 C.J.GREEN, B.S.VOLD (1983) NUCL. ACIDS RES. 11, 5763-5774
Y.YAMADA,R.OHNT,M.ISHIKURA(1983) NUCL.ACIDS RES. 11, 3037-3045
1149 F.J.DEL REY,T.F.DONAHUE,G.R.FINK(1982) PROC.NATL.ACAD.SCI.U.S.A. 79, 4138-4142
1150 V.GAMULIN,J.RAO,B.APPEL,S,BURNER-SMITH,F.YAMAO,D.SOELL(1983) NUCL.ACIDS RES. 11, 4537-4546
1151 D.DEFRANCO, G.SCHMIDT, D.SOELL (1980) PROC. NATL. ACAD. SCI. USA 77, 3363-3368
B.HOVEMANN, S.SHARP, H.YAMADA, D.SOELL (1980) CELL 19, 889-895
P.H.YEN,N.DAVIDSON(1980) CELL 22, 137-148



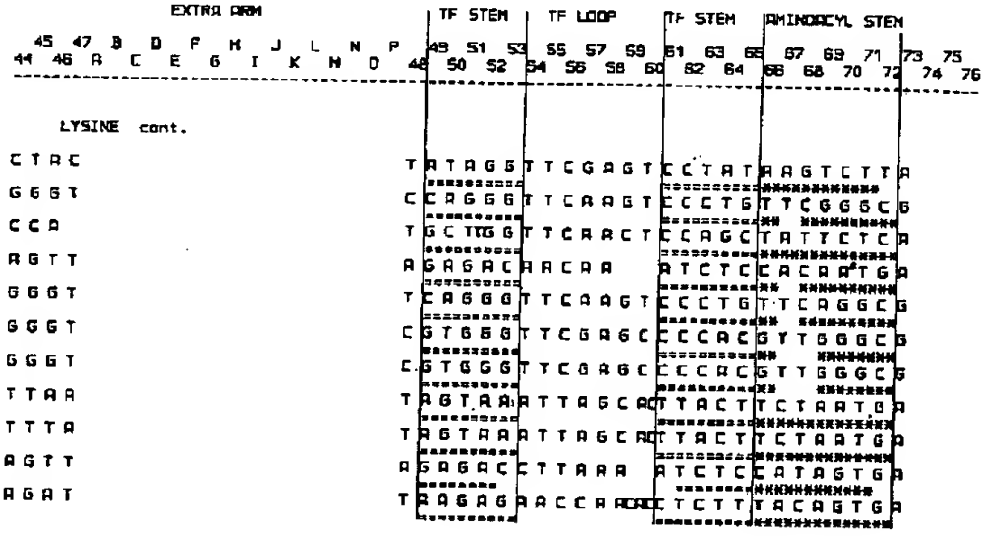
1096/0 ALIGNMENT IS ARBITRARY
1097/0 ALIGNMENT IS ARBITRARY
1099/0 ALIGNMENT IS ARBITRARY

1149/38 AFTER RESIDUE 38 INTERVENING SEQUENCE
1150/37 AFTER RESIDUE 37 INTERVENING SEQUENCE

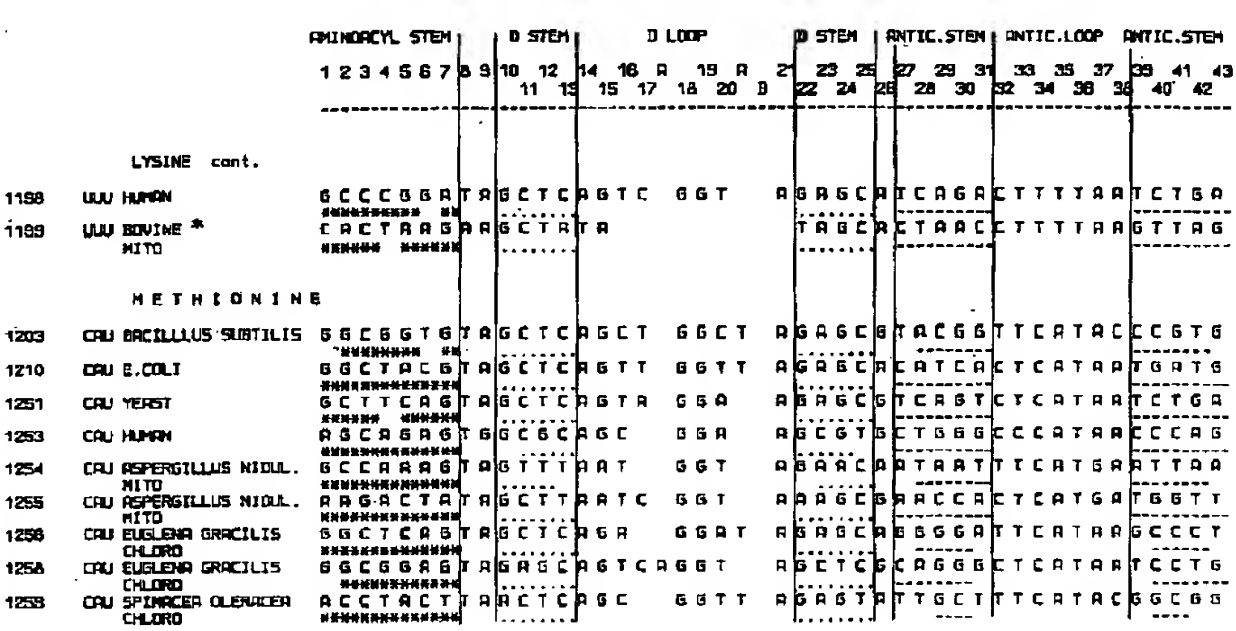
Nucleic Acids Research



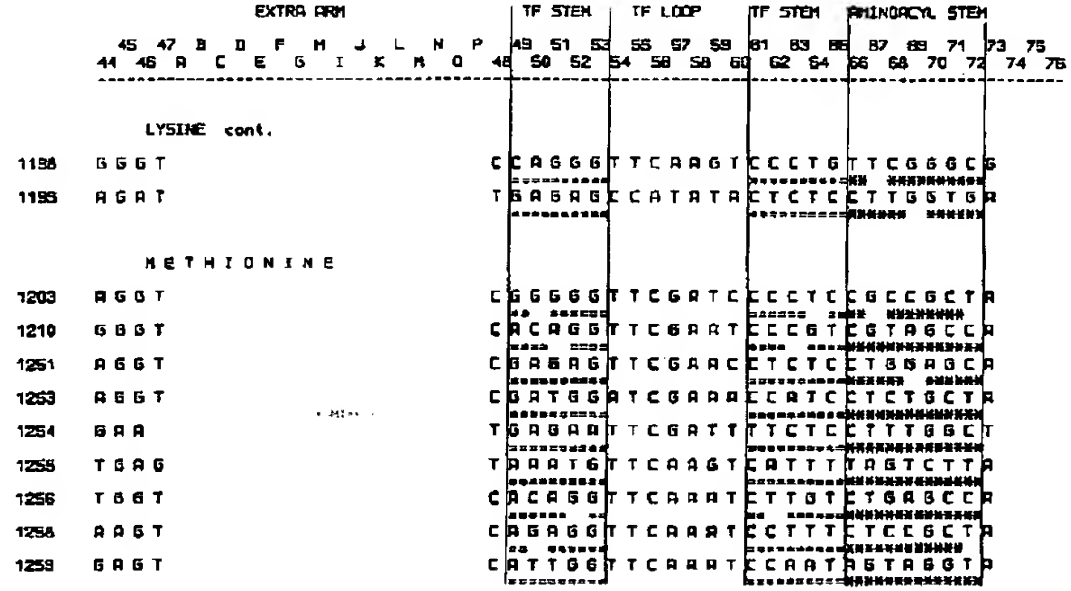
1152 M.G.KOECHEL, C.A.LAZARUS, H.BASAK, H.KUENTZEL (1981) CELL 23, 625-633
1153 D.DEFRANCO, R.B.BURKE, S.KAYASNI, G.A.TENER, R.C.MILLER, JR., O.SOELL (1982) NUCL. ACIDS RES. 10, 5799-5808
1154 S.B.BOHNIZ, A.TZAGOLOFF (1980) J. BIOL. CHEM. 255, 9075-9081
D.L.MILLER, A.C.MARTIN (1981) CURR. GEN. 4, 135-143
1155 P.CANTATORE, C.DE BENEDETTO, G.GADALETA, A.GALLERANI, A.A.KROON, R.HOLTROP, C.LANAVE, G.PEPE, C.QUAGLIARIELLO, C.SACCONE, E.SBISA (1982) NUCL. ACIDS RES. 10, 3279-3289
1156 J.H.HAM, J.O.HARDING (1983) NUCL. ACIDS RES. (1.
2033-2044
1157 F.SEKIYA, I.NISHIZAWA, K.OATSUDA, Y.TAYA, S.NISHIMURA (1982) NUCL. ACIDS RES. 10, 6431-6439
1158 E.WITTIG, B.WITTIG (1982) NATURE 297, 31-38
1185 B.H.L.DE BRUIN(1983) NATURE 304, 234-241
1186 D.D.CLARY, D.R. WOLSTENHOLME (1983) NUCL. ACIDS RES. 11, 4211-4227
1196 A.J.BIBB, R.A.VAN ETTEN, C.T.WRIGHT, M.W.WALBERG, C.A.CLAYTON (1981) CELL 26, 167-180
1197 S.ANDERSON ET AL. (1981) NATURE 290, 457-465



1155/0 ALIGNMENT IS ARBITRARY
1155/70 IS REPORTED TO BE T IN R. GROSSKOPF, H.FELDMANN (1981) CURR. GEN. 4, 151-158
1159/0 ALIGNMENT IS ARBITRARY
1186/0 ALIGNMENT IS ARBITRARY
1196/0 ALIGNMENT IS ARBITRARY
1197/0 ALIGNMENT IS ARBITRARY

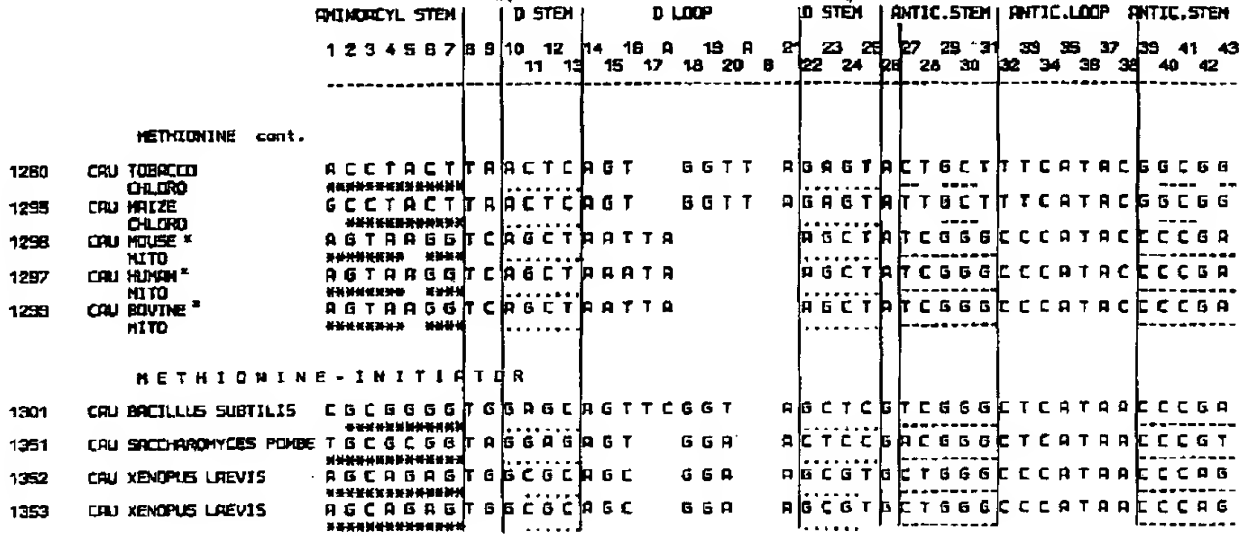


1198 R.L.ROY, M.COUXIE, R.JUCKLAND (1982) NUCL. ACIDS RES. 10, 7313-7322
1199 S.ANDERSON, M.H.L.DEBRUIJN, A.R.COULSON, I.C.EPERON, F.SANGER, I.G.YOUNG (1982) J. MOL. BIOL. 156, 683-717
1203 C.J.GREEN, B.S.VOLD (1983) NUCL. ACIDS RES. 11, 5763-5774
1210 H.KAKAJIMA, H.OZERI, Y.SHIBUYA (1981) CELL 23, 239-247
1251 J.OLAN, H.FELDMANN (1980) NUCL. ACIDS RES. 8, 1775-1786
1253 T.SANTOS, R.ZABLOFF (1981) CELL 23, 699-709
1254 H.G.KUECHEL, C.R.LAZARUS, R.BABAK, R.KUENTZEL (1981) CELL 23, 625-633
1255 M.G.KOECHEL, C.R.LAZARUS, R.BABAK, H.KUENTZEL (1981) CELL 23, 625-633
1256 R.J. KOLLINGSWORTH, R.B. HALLICK (1982) J. BIOL. CHEM. 257, 12795-12799
1258 G.G.ENBASIN, R.B.HALLICK (1983) J.BIOL.CHEM. 258, 5512-5518
1259 W.GRUISSEM, B.R.GREENBERG, G.ZURAWSKI, D.R.PRESCOTT, R.B.HALLICK (1983) CELL 35, 815-826

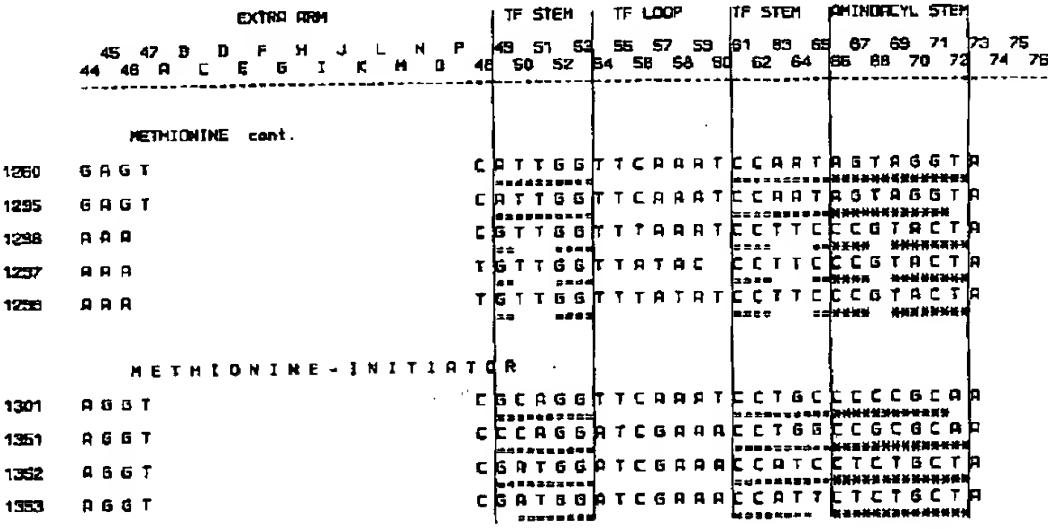


.1199/0 ALIGNMENT IS ARBITRARY

Nucleic Acids Research

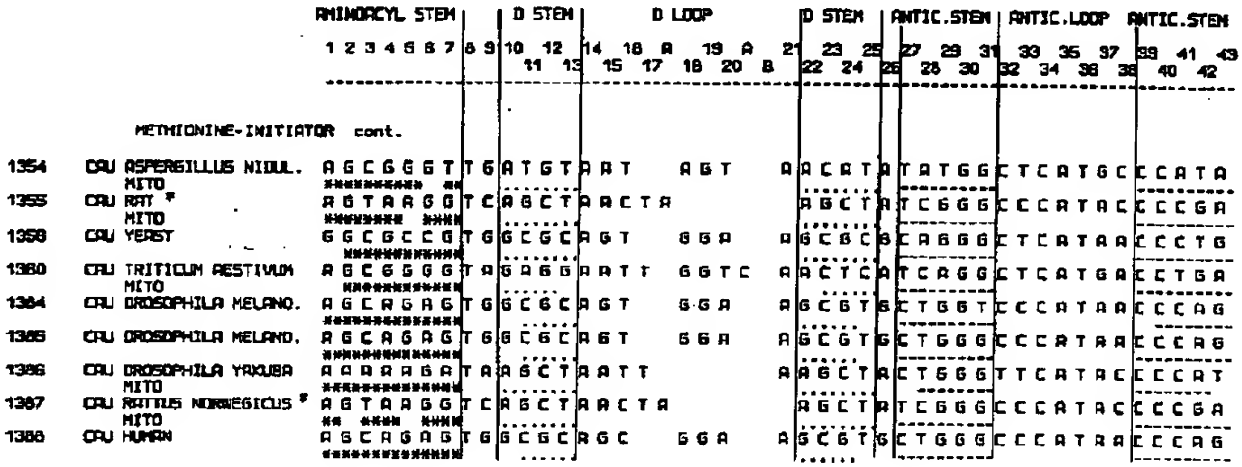


1280 H.BERG, A.KATO, K.SHINOZAKI, M.SUGIURA (1982) NUCL. ACIDS RES. 10, 7511-7520
1295 A.A.STEINMETZ, E.T.KREBBERS, Z.SCHWARZ, E.J.GUBBINS, L.BOGORAD (1983) J.BIOL.CHEM. 258, 5503-5511
1298 R.J.DIBB, R.A.VAN ETTEN, C.T.WRIGHT, M.W.WALBERG, D.A.CLAYTON (1981) CELL 26, 167-180
1297 S.ANDERSON (1981) NATURE 290, 457-465
1299 S.ANDERSON, M.H.L.DEBRUIJN, A.R.COULSON, I.C.EPERON, F.SANGER, I.G.YOUNG (1982) J. MOL. BIOL. 156, 683-717
1301 C.J.GREEN, B.S.VOLD (1983) NUCL. ACIDS RES. 11, 5763-5774
1351 J.RAO, D.SCHMIDT, D.SOELL (1980) CELL 21, 509-516
H.HOTTINGER,D.PEARSON,F.YA RAO,U.BARNLIN,L.COOLEY,T.COOPER D.SOELL(1982) MOL.GEN.GENET. 188, 239-234
1352 J.L.TELFORD, A.KRESSMANN, R.A. KOSKI, R.GROSSCHEDL, F.MUELLER, G.G. CLARKSON, M.L. BIRNSTIEL (1979) PROC. NATL. ACAD. SCI. USA 76, 2590-2594
M.R.FOLK,M.HOFFSETTER(1983) CELL 33, 583-593
R.A.KOSKI,B.G.CLARKSON(1982) J.BIOL.CHEM. 257, 4514-4521
1353 R.A.HIPSKIND,S.G.CLARKSON(1983) CELL 34, 881-890
R.A.HIPSKIND,A.RAZABBAND,J.CORLEY-S.G.CLARKSON(1982: COLD S HARB 47, 873-878

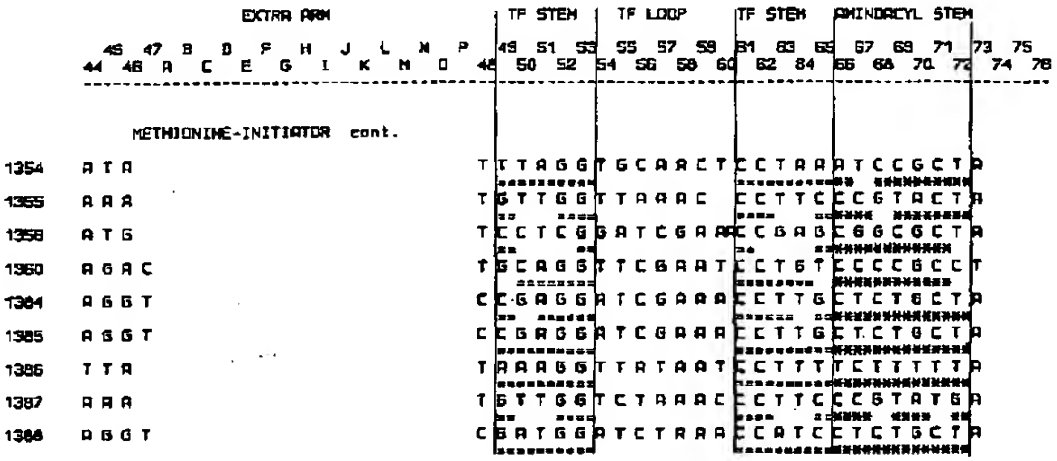


1296/0 ALIGNMENT IS ARBITRARY
1297/0 ALIGNMENT IS ARBITRARY

1299/0 ALIGNMENT IS ARBITRARY

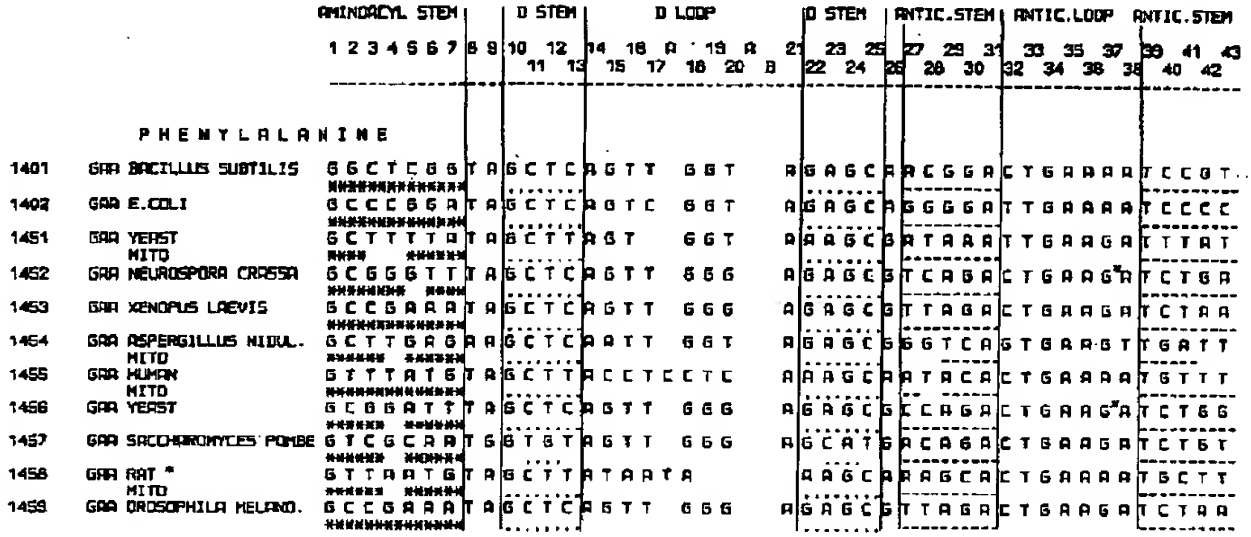


1354 H.B.KOECHEL, C.M.LAZARUS, N.BAŞAK, N.KUENTZEL (1981) CELL 23, 625-633
1355 .CANTATORE, C.DE BENEDETTO, A.BADALETA, R.GALLERANI, A.M.KROON, .HOLTROP, C.LAHAVE, G.PEPE, C.QUAGLIARIELLO, A.SACCONE, E.SBISA 1982) NUCL. ACIDS RES. 10, 3279-3289
1358 A.VENEGAS, E.GONZALEZ, P.BULL, P.VALENZUELA (1982) NUCL. ACIDS RES. 10, 1093-1096
1360 A.W.GRAY,D.F.SPENCER(1983) FEBS LETTERS 161, 323-327
1364 S.SHARP, D.DEFRANCO, A.SILBERKLANG, M.A.MOSBACH, T.SCHMIDT, E.KUBLI, J.P.GERGEN, P.C.WENSINK, D.SOELL (1981) NUCL. ACIDS RES. 9, 5867-5882
1365 S.SHARP, D.DEFRANCO, A.SILBERKLANG, M.A.MOSBACH, T.SCHMIDT, E.KUBLI, J.P.GERGEN, P.C. WENSINK, D.SOELL (1981) NUCL. ACIDS RES. 9, 5867-5882
1366 D.O.CLARY, J.A.GODDARD, S.C.MARTIN, C.M.-R.FAURON, D.R.WOLSTENHOLME NUCL. ACIDS RES. 10, 6419-6427
1367 D.R.WOLSTENHOLME, C.M.-R.FAURON, J.A.GODDARD (1982) GENE 20, 63-69
1368 T.SANTOS, A.ZABLOFF (1981) CELL 23, 699-709

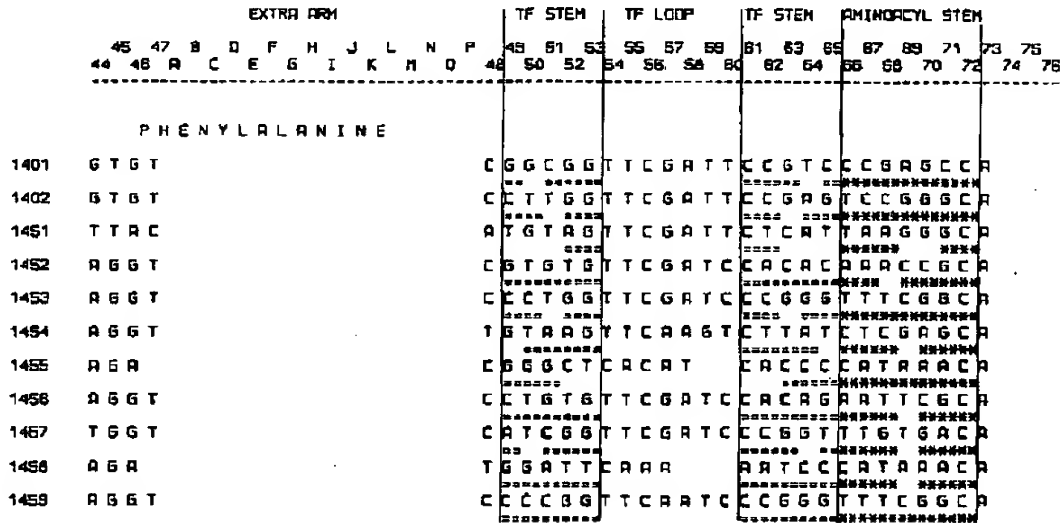


1335/0 ALIGNMENT IS ARBITRARY

1367/0 ALIGNMENT IS ARBITRARY

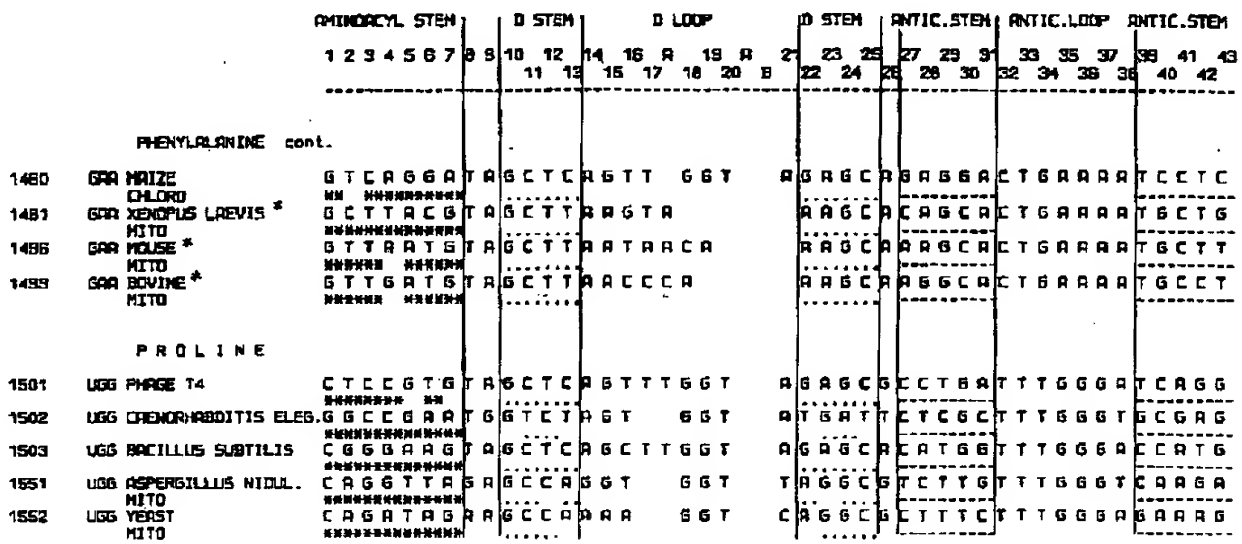


1401 C.J.GREEN, B.S.VOLD (1983) NUCL. ACIDS RES. 11, 5763-5774
Y.YAMADA,M.OHNI,H.ISHIKURA(1983) NUCL.ACIDS RES. 11, 3037-3045
1402 I.SCHWARTZ,R.-A.KLOTZKY,O.ELSEVIERS,P.J.GALLAGHER,M.KRAUSKOPF,H.A.O.
SIDDIQUI,J.F.W.WONG,S.A.ROE(1983) NUCL.ACIDS RES. 11, 4379-4389
1451 D.L.MILLER, W.C.MARTIN, H.B.PHAM, J.E.DONELSON (1980) J.BIOL.
CHEM. 254, 11735-11740
1452 E.SELKER, C.YANOFSKY (1980) NUCL. ACIDS RES. 8, 1033-1042
1453 F.MUELLER, B.G.CLARKSON (1980) CELL 19, 345-353
1454 H.G.KOECHEL, C.M.LAZARUS, W.BARAK, H.KUENTZEL (1981) CELL 24,
625-633
1455 ANDERSON ET AL. (1981) NATURE 290, 457-465
S.CREWS,G.ATTARDI(1980) CELL 19, 775-784
1456 P.VALENZUELA, A.VENEGAS, F.WEINBERG, R.BISHOP, W.J.ROTTER (1978)
PROC. NATL. ACAD. SCI. USA 75, 190-194
1457 V.GARDLIN,J.RAO,B.APPEL,N.SUMMER-SMITH,F.YAMAO,B.SOELL(1983) NUCL.
ACIDS RES. 11, 8537-8546
1458 R.KOBAYASHI, T.SEKI, K.YASUMURA, K.KOIKE (1981) GENE 16, 297-307
R.G.GROSSHOFF,H.FELDMANN(1981) CURR.GEN. 4, 191-196
1459 W.R.ADDISON, C.B.ASTELL, A.D.DELANEY, I.C.GILLAN, S.HAYASHI,
A.C.MILLER, B.RAJPUT, R.SMITH, D.G.TAYLOR, G.M.TENER (1982)
J. BIOL. CHEM. 257, 470-473



1452/37 AFTER RESIDUE 37 INTERVENING SEQUENCE
1456/37 AFTER RESIDUE 38 INTERVENING SEQUENCE

1458/0 ALIGNMENT IS ARBITRARY



1460 A.A.STEINMETZ, E.T.KAEBERLE, Z.SCHWARZ, E.J.GUBBINS, L.BOGORAD (1983) J.BIOL.CHEM. 258, 5303-5311

1461 J.F.H.WONG, D.P.MA, B.K.WILSON, B.A.ROE (1982) NUCL. ACIDS RES. 11, 4977-5004

1496 M.J.BIBB, R.A.VAN ETTEN, C.T.WRIGHT, M.W.WALBERG, D.A.CLAYTON (1981) CELL 26, 167-180

1499 S.ANDERSON, M.H.L.DEBRUIJN, A.R.COULSON, I.C.EPERON, F.SANGER, I.G.YOUNG (1982) J. MOL. BIOL. 156, 683-717

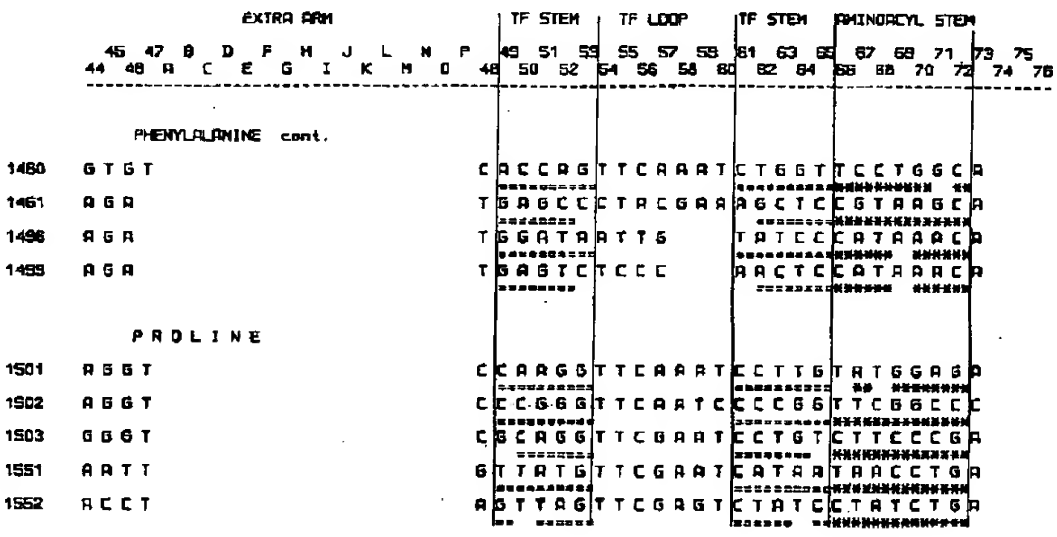
1501 K.FUKADA, J.ABELSON (1980) J.MOL.BIOL. 139, 377-391

1502 G.CILIBERTO, L.CASTAGNOLI, D.A.MELTON, R.CORTESE (1982) PROC. NATL. ACAD. SCI. USA 79, 1195-1199
T.A.TRANGUILA, R.CORTESE, D.MELTON, J.D.SMITH, NUCL. ACIDS RES. 10, 7919-7943

1503 C.J.GREEN, B.S.VOLD (1983) NUCL. ACIDS RES. 11, 5763-5774;
E.F.WASMUTH, J.N.HANSEN (1983) J. BIOL. CHEM. 258, 293-298

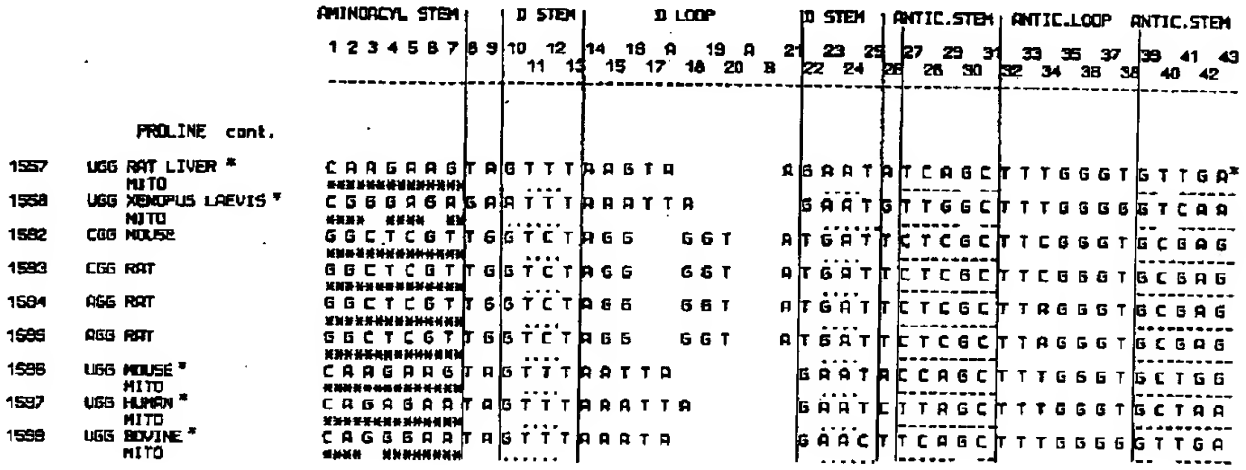
1551 H.B.KOECHEL, C.R.LAZARUS, W.BASKE, H.KUENTZEL (1981) CELL 23, 625-633

1552 D.NEWMAN, M.D. PHAM, K. UNDERBRINK-LYON, N.C.MARTIN (1980) NUCL. ACIDS RES. 8, 5007-5016; (1981) NUCL. ACIDS RES. 9, 2421

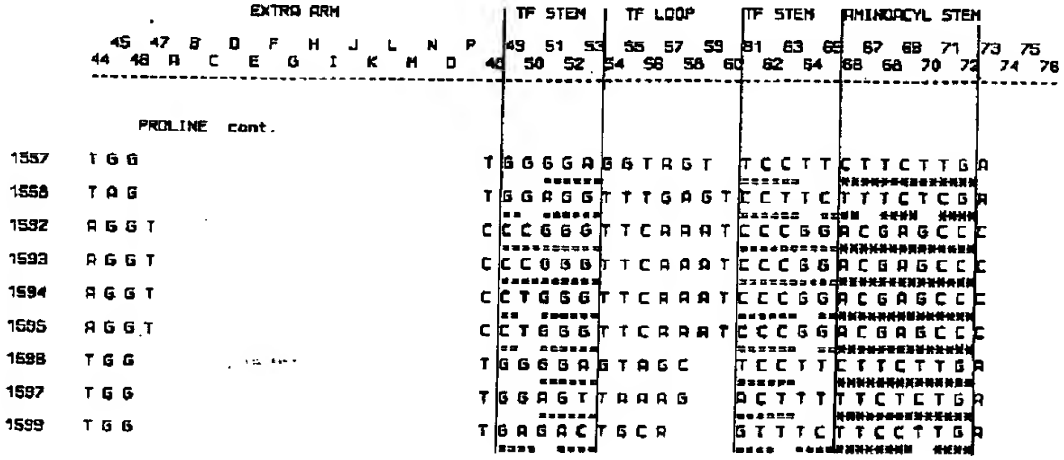


1461/D ALIGNMENT IS ARBITRARY
1496/D ALIGNMENT IS ARBITRARY

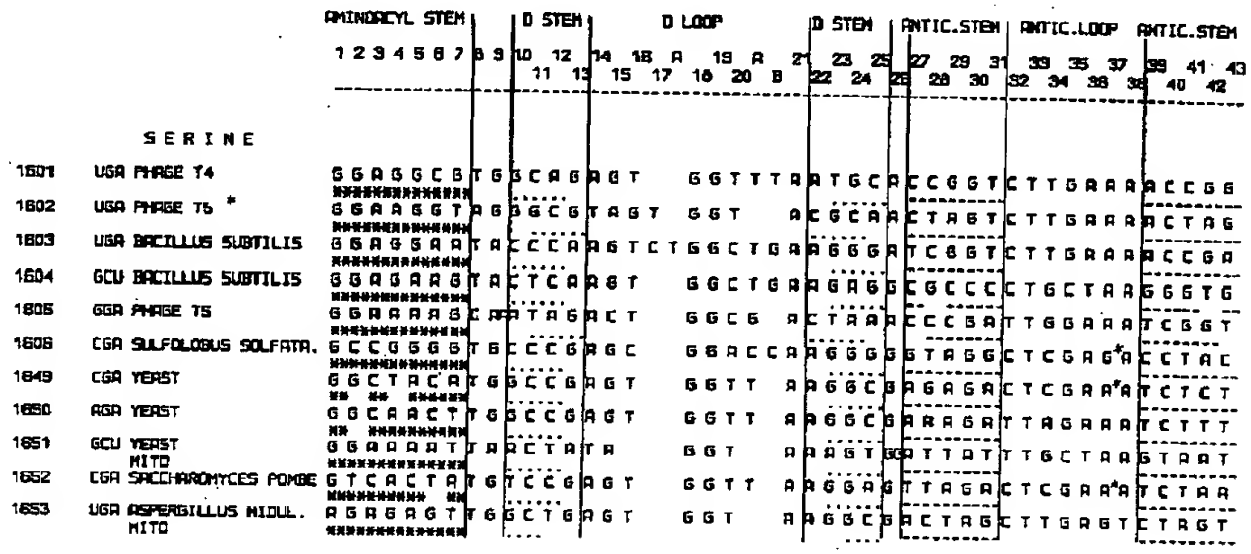
1499/D ALIGNMENT IS ARBITRARY



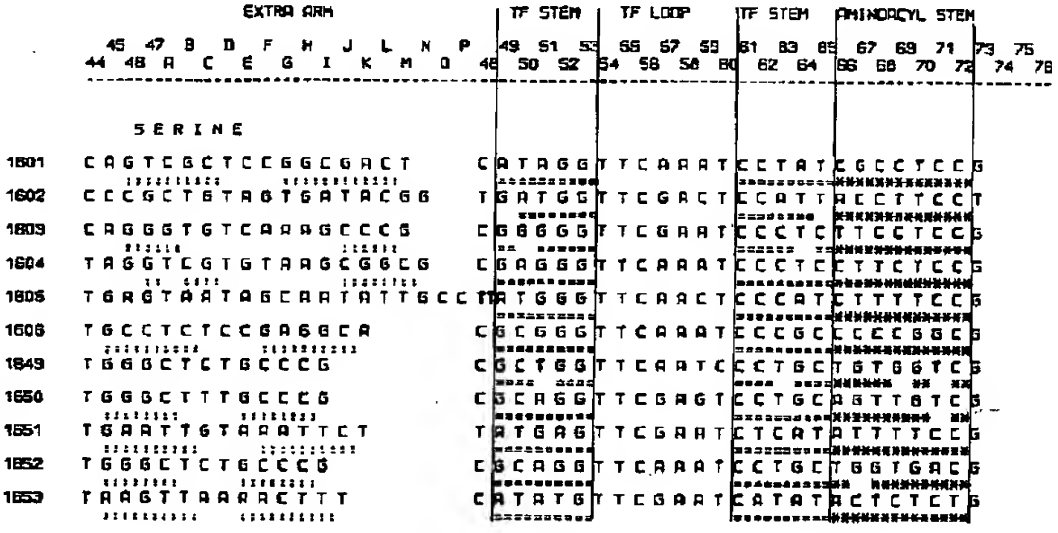
1557 K.KOIKE,H.KOBAYASHI,K.YAGIHURA,A.TAIRA,E.YOSHIDA,A.INAI(1982) GENE 20, 177-185
1558 J.F.H.OHNO, D.F.MA, R.K.WILSON, B.A.RDE (1983) NUCL. ACIDS RES. 11, 4977-5004
1582 J.C.HU,R.D.COTE,E.LUND,J.E.DALBERG(1983) NUCL.ACIDS RES. 11,4809-4821
1583 T.SEKIYA, R.NISHIZAWA, K.MATSUDA, Y.TATA, S.NISHIMURA (1982) NUCL. ACIDS RES. 10, 6411-6419
1584 T.SEKIYA, R.NISHIZAWA, K.MATSUDA, Y.TATA, S.NISHIMURA (1982) NUCL. ACIDS RES. 10, 6411-6419
1585 T.SEKIYA, R.NISHIZAWA, K.MATSUDA, Y.TATA. S.NISHIMURA (1982) NUCL. ACIDS RES. 10, 6411-6419
1586 R.J.BIBB, R.A.VAN ETTEN, C.T.WRIGHT, M.W.WALBERG, D.A.CLAYTON (1981) CELL 26, 167-180
1587 S.ANDERSON ET AL. (1981) NATURE 290, 457-465
1588 S.ANDERSON, R.H.L.DEBRUIJN, A.R.COULSON, I.C.EPERON, F.SANGER, I.G.YOUNG (1982) J.MOL. BIOL. 156, 683-717



1557/0 ALIGNMENT IS ARBITRARY
1557/42 IS REPORTED TO BE G-IN O.QUERTZ,A.FELDMANN(1982) CURR.GEN. 5, 221-225
1558/0 ALIGNMENT IS ARBITRARY
1582/0 ALIGNMENT IS ARBITRARY
1583/0 ALIGNMENT IS ARBITRARY
1584/0 ALIGNMENT IS ARBITRARY
1585/0 ALIGNMENT IS ARBITRARY
1586/0 ALIGNMENT IS ARBITRARY
1587/0 ALIGNMENT IS ARBITRARY
1588/0 ALIGNMENT IS ARBITRARY

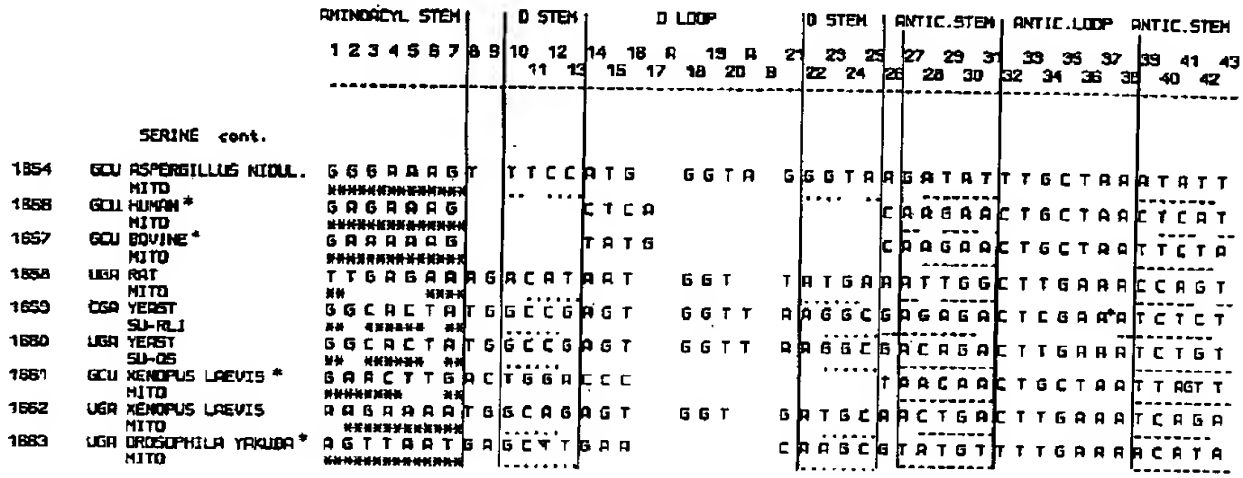


1601 K.FURADA, J.ABELSON (1980) J.MOL.BIOL. 139, 377-391
E.P.RAZZARA,O.PUNKETT,M.H.RECLAIM(1981) PROC.NATL.ACAD.SCI.U.S.A. 78, 689-692
1602 V.A.KRYUKOV,V.N.KSENZENKO,A.V.KALINAN,A.A.BAYEV(1983) FEBS LETTERS 156,123-127
1603 C.J.GREEN, B.B.VOLD (1983) NUCL. ACIDS RES. 11, 5763-5774
1604 C.J.GREEN, B.S.VOLD (1983) NUCL. ACIDS RES. 11, 5763-5774
1605 V.A.KRYUKOV,V.N.KSENZENKO,A.V.KALINAN,A.A.BAYEV(1983) FEBS LETTERS 153,123-127
1606 B.P.KAINE, R.GUPTA, C.R.WOESE (1983) PROC. NATL. ACAD. SCI. USA 80, 3309-3311
1649 R.C.OGDEN ET AL.(1980) IN: TRANSFER RNA,D.SOELL,J.ABELSON,P.SCHIMMEL COLD SPRING HARB.. 173-179
1650 G.G.PAGE, S.O.HALL (1981) NUCL. ACIDS RES. 9, 921-934;
R.E.BAKER, A.EIGEL, D.VOLOTEL, N.FELDMARN (1982) EMBO JOURNAL 1, 291-295
1651 S.G.BONITZ. A.TZAGOLOFF (1980) J.BIOL.CHEM. 255, 9075-9081
D.L.MILLER, N.C.MARTIN (1981) CURR. GEN. 4, 135-143
1652 J.RAO, D.SCHMIDT, D.SOELL (1980) CELL 21, 509-516
1653 H.G.KOECHEL, C.J.LAZARUS. N.BASAK, M.KUENTZEL (1981) CELL 23, 625-633

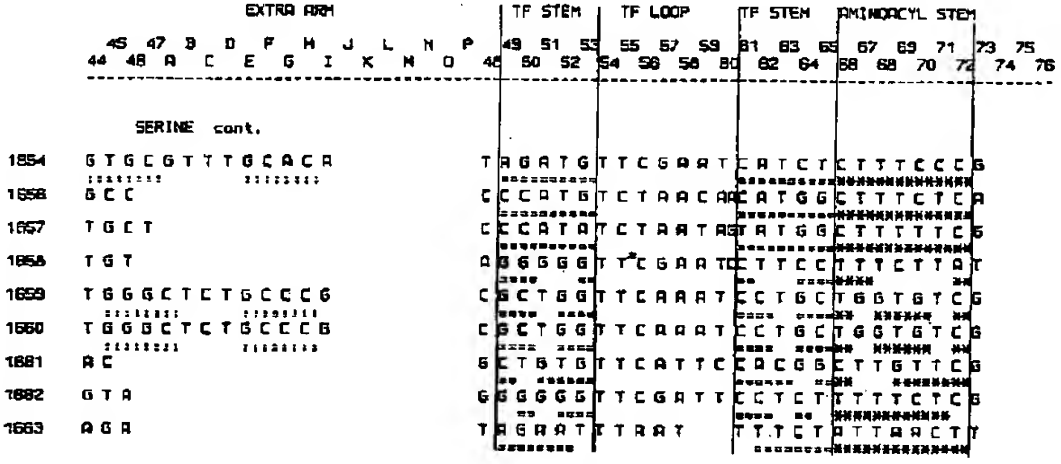


1602/0 IN KRYUKOV ET.AL. ALIGNMENT DEVIATION
1606/37 POSSIBLE SITE OF INTRON CLEAVAGE

1649/37 AFTER RESIDUE 37 INTERVENING SEQUENCE
1652/37 AFTER RESIDUE 37 INTERVENING SEQUENCE

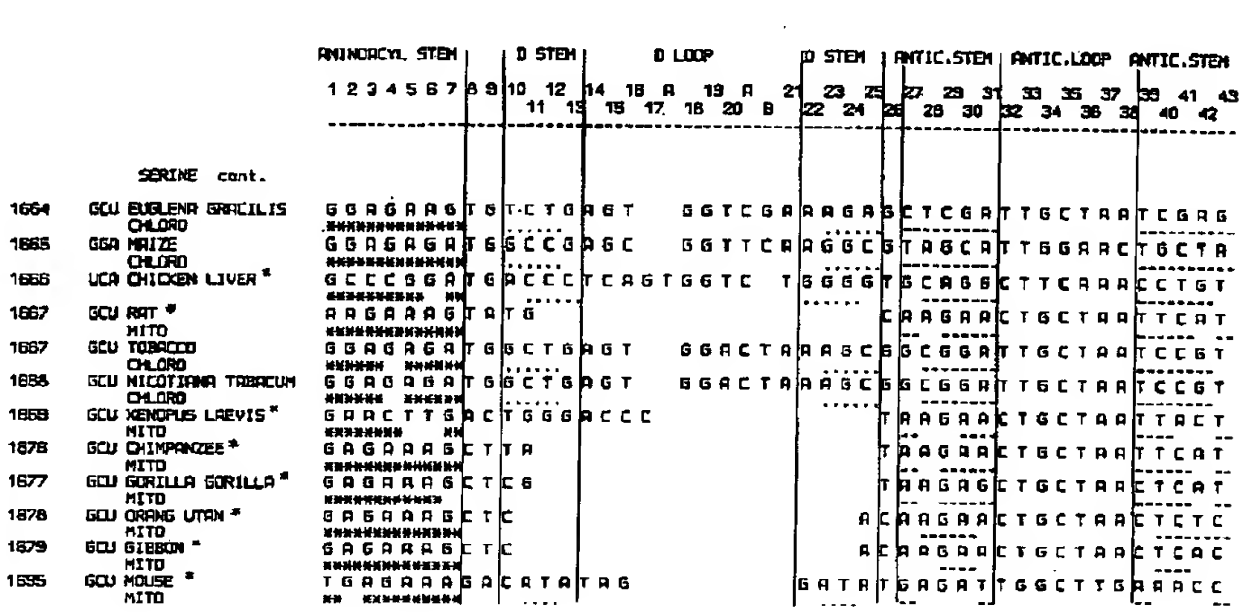


1654 H.B.KOECHEL, C.R.LAZARUS, R.WASM, N.KUENTZEL (1981) CELL 23, 625-633
1656 ANDERSON ET AL. (1981) NATURE 290, 457-465
U.E.BROWN,E.H.PRAGER,A.WANG,A.C.WILSON(1982) J.MOL.EVOL. 18, 225-239
1657 A.M.L. DE BRUIJN, P.H. SCHREIER, I.C. EPERON, B.G.BARRELL, E.Y.CHEN, P.S. ARMSTRONG, J.F.H.DOWD, B.A.ROE (1980) NUCL. ACIDS RES. 8, 3213-3222
S.ANDERSON, A.M.L.DEBRUIJN, A.R.COULSON, I.C.EPERON, F.SANGER, I.G.YOUNG (1982) J. MOL. BIOL. 156, 683-717
1658 R.GROSSKOPF,H.FELDMANN(1981) CURR.GEN. 4, 151-156
1659 A.V.OLSON, G.S.PAGE, A.SENTENAC, P.U.PIPER, A.WORTHINGTON, A.B.WEISS, B.D.HALL (1981) NATURE 291, 464-467
1660 A.V.OLSEN,G.S.PAGE,A.SENTENAC,P.U.PIPER,A.WORTHINGTON,A.B.WEISS, B.D.HALL(1981) NATURE 291, 464-467
J.R.BROACH,L.FRIEDMAN,F.SHERMAN (1981) J.MOL.BIOL. 150, 375-387
S.R.SANDREYER,A.V.OLSON (1982) PROC.NATL.ACAD.SCI.USA 79, 7674-7678
1661 B.A.ROE, J.F.H.WONG, DIN-POW MA, R.K.WILSON (1983) PERSONAL COMMUNICATION
1662 B.A.ROE, J.F.H.WONG, DIN-POW MA, R. WILSON (1983) PERSONAL COMMUNICATION
1663 D.O.CLARY, J.A.WAHLEITHNER, D.A.WOLSTENHOLME (1983) NUCL. ACIDS RES. 11, 2411-2425

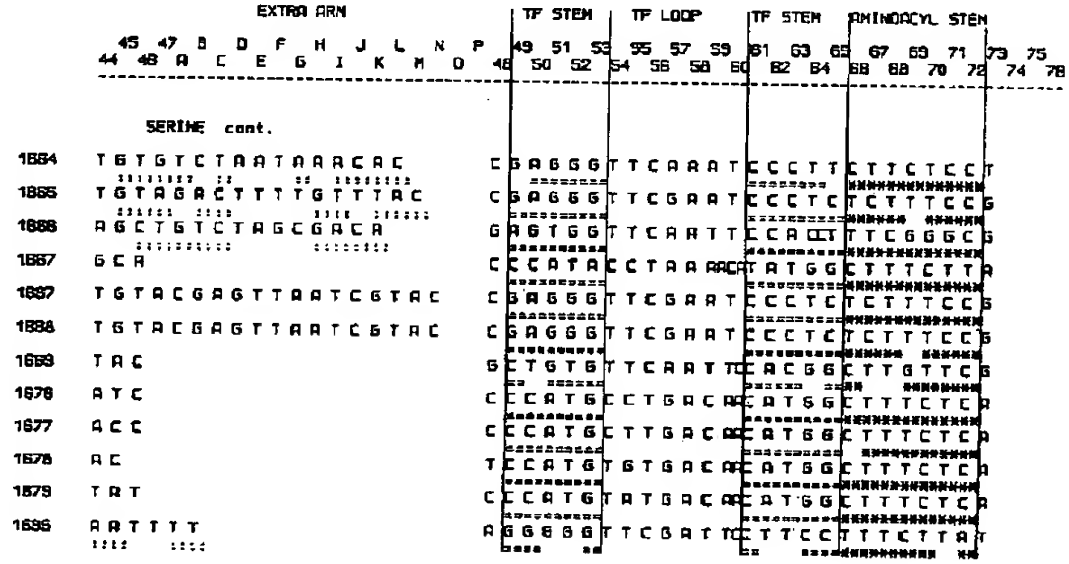


1656/0 ALIGNMENT IS ARBITRARY
1657/0 ALIGNMENT IS ARBITRARY
1658/55 T WAS NOT FOUND BY P.CANTATORE, C.DE BENEDETTO, G.GADALETA, R.GALLERANI, A.A.KROON, A.MOLTROP, C.LANAVE, G.PEPE, C.SBISA, E.SBISA(1982) NUCL. ACIDS RES. 10, 3279-3289
1659/37 AFTER RESIDUE 37 INTERVENING SEQUENCE
1661/0 ALIGNMENT IS ARBITRARY
1663/0 ALIGNMENT IS ARBITRARY

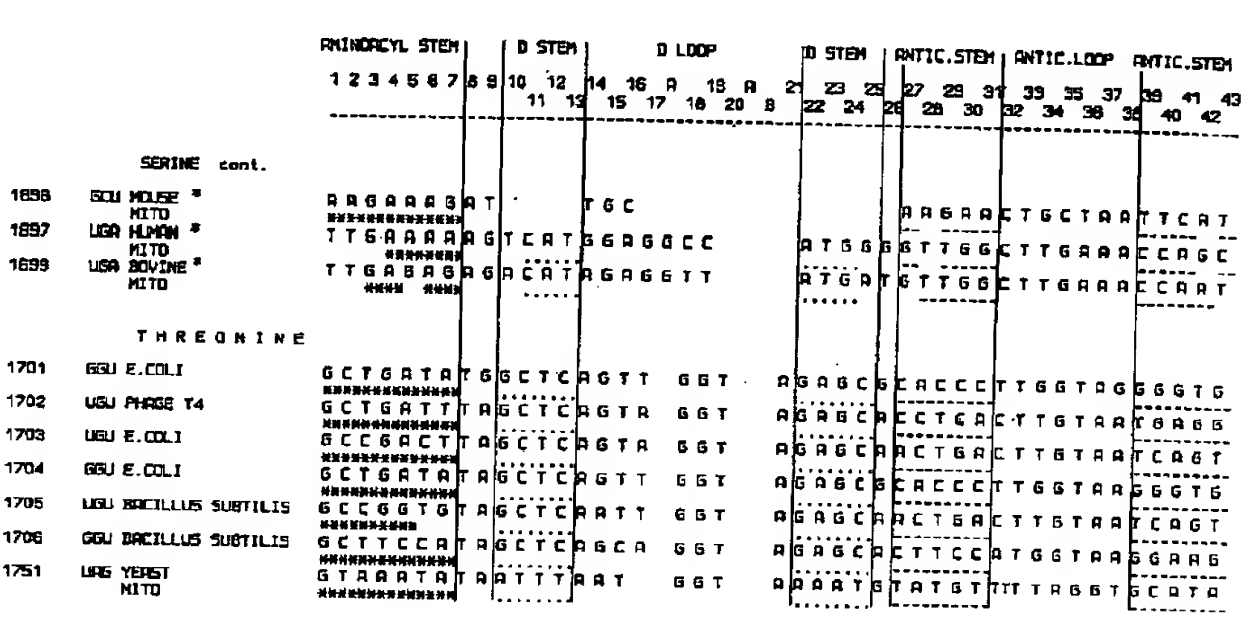
Nucleic Acids Research



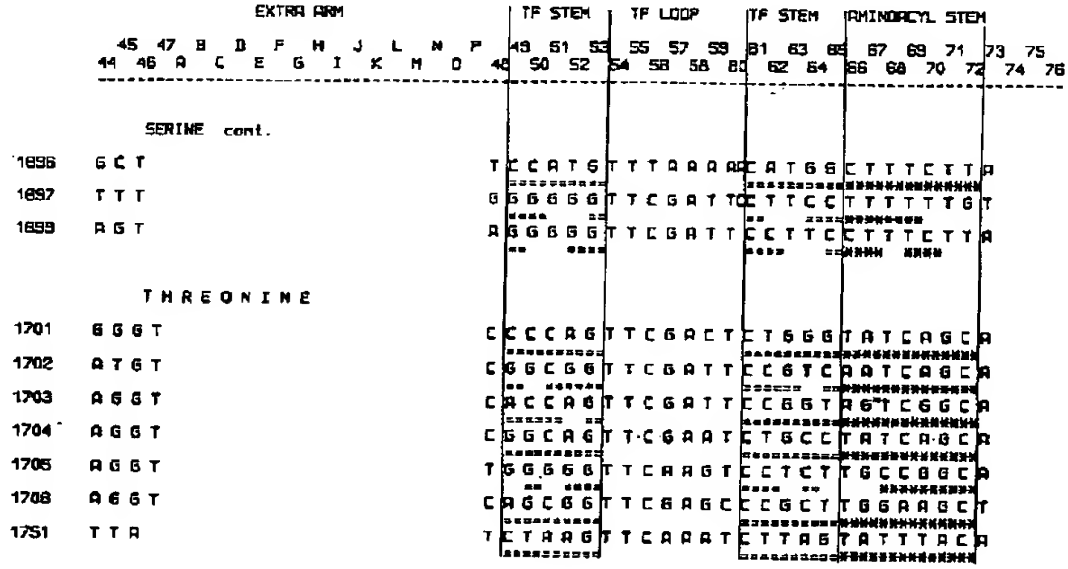
1664 G.D.KARAITH, R.B.HALLICK (1983) J. BIOL. CHEM. 258, 5312-5316
1665 A.A.STEINMETZ, E.T.KREBBERS, Z.SCHWARZ, E.J.GUBBINS, L.BOGORAD (1983) J.BIOL.CHEM. 258, 5503-5511
1666 D.L.HATFIELD, R.S.GUDDOCK, F.C.EDEN (1983) PROC. NATL. ACAD. SCI. U.S.A. 80, 4940-4944
1667 R.B.GROSSKOPF,H.FELDMANN(1981) CURR.GEN. 4, 191-196
1667 H.DENO,A.SUGIURA(1983) NUCLEIC ACIDS RES. 23, 8407-8414
1668 H.DENO,A.SUGIURA(1983) NUCL.ACIDS RES. 11, 8407-8414
1669 J.F.H.WONG,D.-F.MA,R.E.WILSON,B.A.ROE(1982) FED.PROC. 42(7), 2240
1676 W.M.BROWN,E.M.PRAGER,A.WANG,A.C.WILSON(1982) J.MOL.EVOL. 18, 225-239
1677 W.M.BROWN,E.M.PRAGER,A.WANG,A.C.WILSON(1982) J.MOL.EVOL. 18, 225-239
1678 W.M.BROWN, E.M.PRAGER, A.WANG. A.C. WILSON (1982) J. MOL. EVOL. 18, 225-239
1679 W.M.BROWN, E.M.PRAGER, A.WANG, A.C.WILSON (1982) J. MOL. EVOL. 18, 225-239
1695 R.J.MEHS, R.A.VAN ETTEN, C.T.WRIGHT, R.H.WALBERG, D.A.CLAYTON (1981) CELL 26, 167-180



1666/0 USA SUPPRESSOR TRNA
1667/0 ALIGNMENT IS ARBITRARY
1669/0 ALIGNMENT IS ARBITRARY
1676/0 ALIGNMENT IS ARBITRARY
1677/0 ALIGNMENT IS ARBITRARY
1678/0 ALIGNMENT IS ARBITRARY
1679/0 ALIGNMENT IS ARBITRARY
1695/0 ALIGNMENT IS ARBITRARY

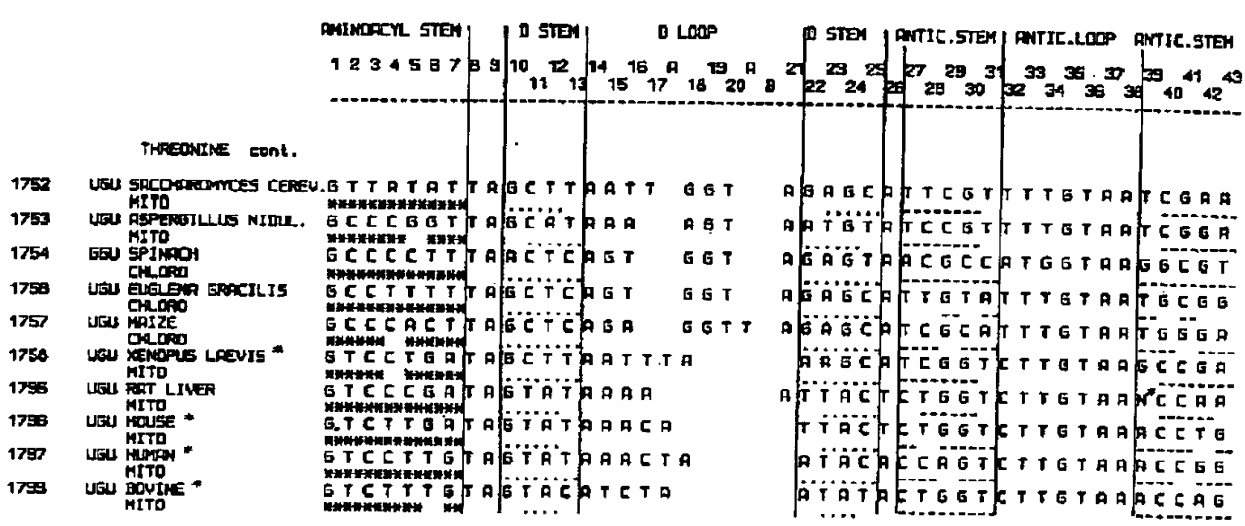


1696 A.J.BIBB, W.A.VAN ETTEN, C.T.WRIGHT, A.W.WALBERG, D.A.CLAYTON (1981) CELL 26, 167-180
1697 S.ANDERSON ET AL. (1981) NATURE 290, 457-465
1699 S.ANDERSON, A.H.L.DEBRUIJN, A.R.COULSON, I.C.EPERON, F.SANGER, I.G.YOUNG (1982) J. MOL. BIOL. 156, 683-717
1701 G.L.BOESTER, W.A.HOLMES (1980) NUCL. ACIDS RES. 8, 3793-3807
1702 K.FUKADA, J.ABELSON (1980) J. MOL. BIOL. 139, 377-391
G.P.MAZZARA,G.PLUNKETT,U.H.RODLAIN(1981) PROC.NATL.ACAD.SCI.U.S.A. 78, 849-592
1703 L.HUDSON, J.ROSSI, A.LANDY (1981) NATURE 294, 422-427
B.AM,J.D.FRIESEN(1980) GENE 12, 33-39
1704 L.HUDSON, J.ROSSI, A.LANDY (1981) NATURE 294, 422-427
A.AIYAJIMA,T.YOKOTA,Y.TAKEBE,A.NAKAMURA,Y.KAZIRO(1983) J.BIOCHEM. 93, 1101-1108
B.AM,J.D.FRIESEN(1980) GENE 12, 33-39
1705 C.J.GREEN, B.S.VOLD (1983) NUCL. ACIDS RES. 11, 5763-5774
1706 E.F.VANROUSEK, J.M.HANSEN (1983) J. BIOL. CHEM. 258, 291-295
1751 M.LI, A.TZAGOLOFF (1979) CELL 18, 47-53

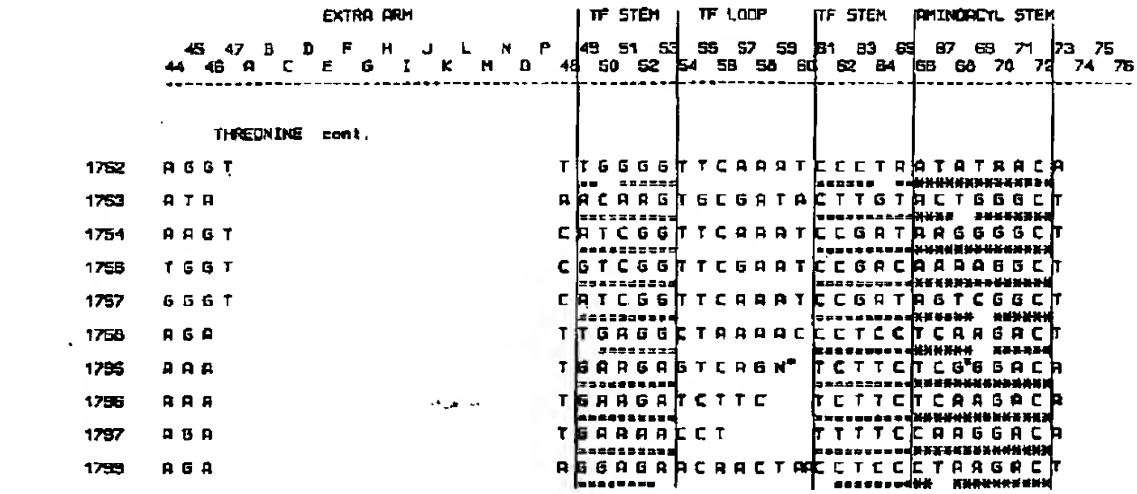


1696/0 ALIGNMENT IS ARBITRARY
1697/0 ALIGNMENT IS ARBITRARY
1699/0 ALIGNMENT IS ARBITRARY

1703/67 IS REPORTED TO BE A C IN A.AIYAJIMA,T.YOKOTA,T.TAKEBE,A.NAKAMURA Y.KAZIRO(1983) J.BIOCHEM. 93, 1101-1108

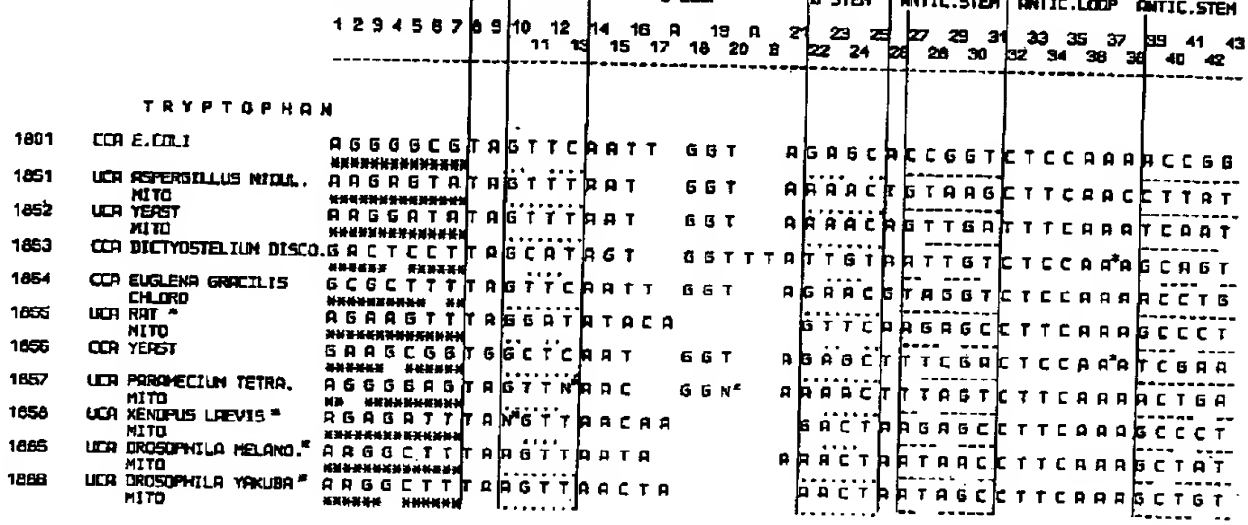


1752 R.E.BERLANI, C.PERTELLA, G.MACINO, A.TZAGOLOFF (1980) J. BACT. 141, 1066-1099
1753 H.U.KOECHEL, C.R, LAZARUS, N.BASAK, H.KUENTZEL (1981) CELL 23, 625-633
1754 R.A.KASHDAN, B.B.DUDOCK (1982) J. BIOL. CHEM. 257, 1114-1116
1756 G.D.MARAGIN, R.D.HALLICK (1983) J.BIOL.CHEM. 258, 5512-5518
1757 A.A.STEINMETZ, E.T.KREDDERS, Z.SCHWARZ, E.J.GIBBINS, L.BOGORAD (1983) J.BIOL.CHEM. 258, 5503-5511
1758 J.F.H.WONG, D.F.RA, R.K.WILSON, B.A.ROE (1983) NUCL. ACIS RES. 11, 4977-5004
1795 G.BÖERTZ,H.FELDMANN(1982) CURR.GEN. 5, 221-225
1796 R.J.GIBB, R.A.VAN ETTEN, C.T.WRIGHT, R.E.WALBERG, D.A.CLAYTON (1981) CELL 26, 167-180
1797 S.ANDERSON ET AL. (1981) NATURE 290, 457-465
1799 S.ANDERSON, M.H.L. DE BRUIJN, A.R.COULSON, I.C.EPERON, F.SANGER, I.G.YOUNG (1982) J. MOL. BIOL. 156, 683-717

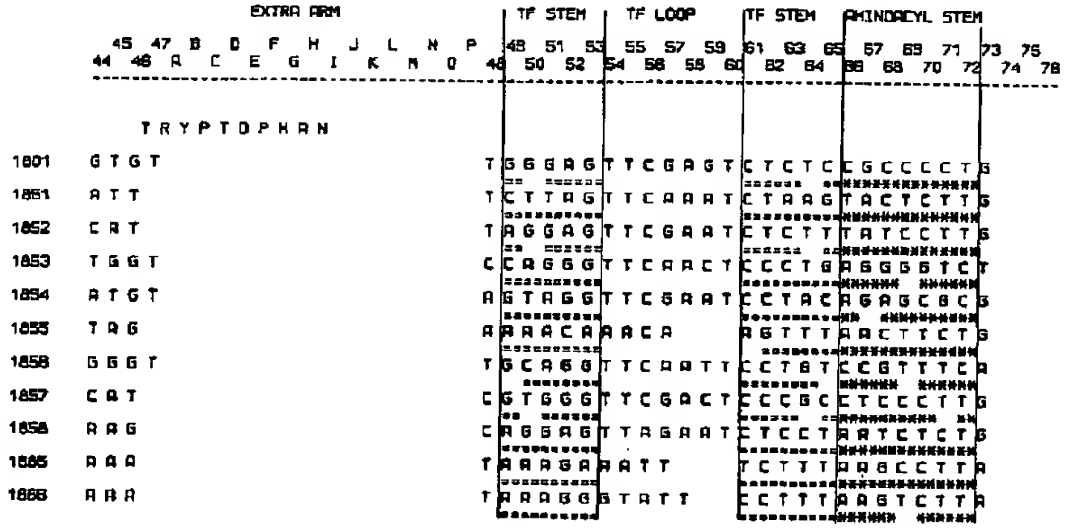


1758/0 ALIGNMENT IS ARBITRARY
1795/39 N IS EITHER D OR A
1795/59 N IS EITHER G OR C
1795/48 IS REPORTED TO BE A IN K.KOISE,H.KUBAYASHI,K.YAGINUMA,M.IAIRA,E.YOSHID R.IBA(1982) GENE 20, 177-185
1796/0 ALIGNMENT IS ARBITRARY
1797/0 ALIGNMENT IS ARBITRARY
1799/0 ALIGNMENT IS ARBITRARY

Nucleic Acids Research



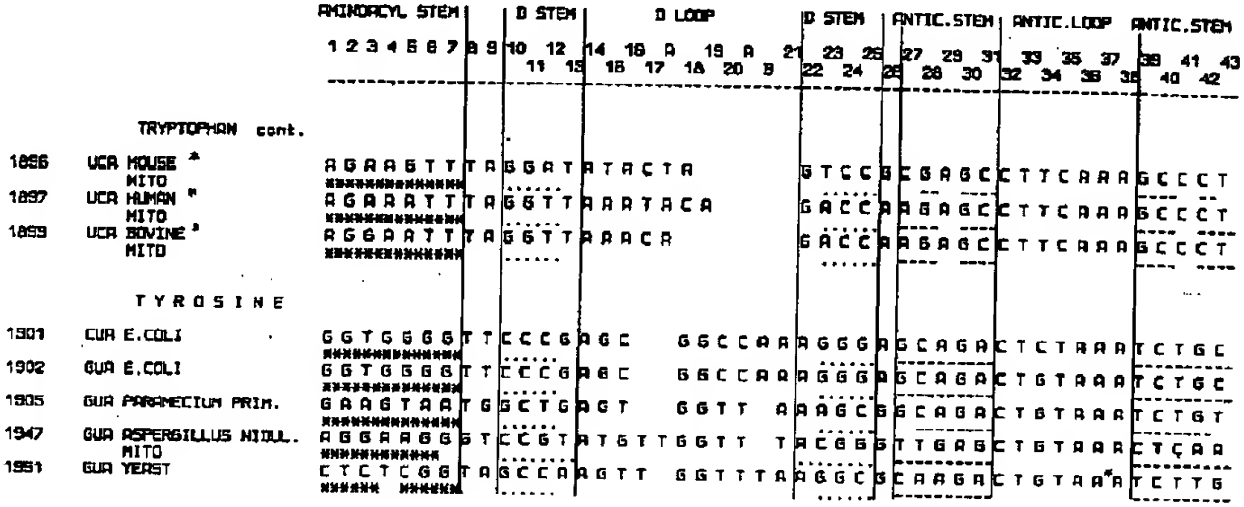
1801 R.A. YOUNG (1979) J.BIOL.CHEM. 254, 12725-12731
1851 H.D.KOECHEL, C.R.LAZARUS, R.BASAK, H.KUENTZEL (1981) GENE 23, 625-633
1852 M.C.MARTIN, H.D.PHAM, K.UNDERBRINK-LYON, D.L.MILLER, J.E.DONELSON (1980) NATURE 285, 579-581
1853 D.M.PEFFLEY, D.L.SOBIR (1981) BIOCHEMISTRY 20, 4015-4021
1854 R.J.HOLLINGSWORTH, R.B.HALLICK (1982) J. BIOL. CHEM. 257, 12795-12799
1855 P.CANTAFORE, C.DE BENEDETTO, D.GADALETA, R.GALLERANI, A.A.KROON, A.HOLTROP, C.LANAVE, G.PEPE, C.QUAGLIARIELLO, C.SACCONE, E.SBISA (1982) NUCL. ACIDS RES. 10, 3279-3289
R.GROSSHOFF, H.FELDMANN (1981) CURR. GEN. 4, 191-196
1856 R.C.GOBEN ET AL.(1980),IN:TRANSFER RNA,D.SOELL,J.ABELSON,P.SCHIMMEL (ED) COLD SPRING HARB.,173-190
1857 J.J.SEILHAMER,D.J.CUMMINGS(1982) MOL.GEN.GENET. 187, 336-339
1858 J.F.H.WONG, D.P.MA, R.K.GILSON, R.A.ROE (1983) NUCL. ACIDS RES. 11, 4977-5004
1885 N.H.L.DE BRUIJN(1983) NATURE 304, 234-241
1886 D.D.CLARY, D.R.WOLSTENHOLME (1983) NUCL. ACIDS RES. 11, 3539-3572



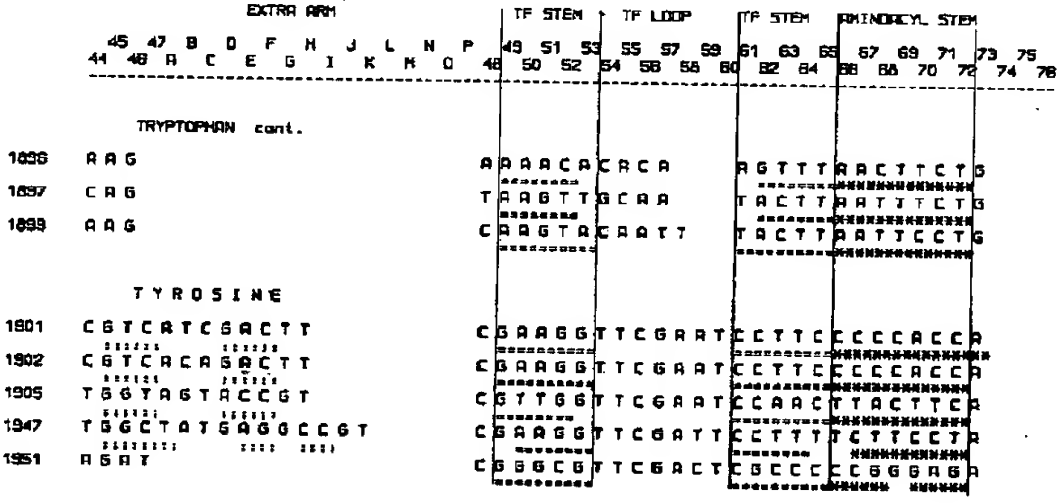
1853/37 AFTER RESIDUE 37 INTERVENING SEQUENCE
1855/0 ALIGNMENT IS ARBITRARY
1856/37 AFTER RESIDUE 37 INTERVENING SEQUENCE
1857/13 R IS EITHER T OR C
1857/20 N IS EITHER T OR A

1858/0 ALIGNMENT IS ARBITRARY
1858/10 N IS EITHER G OR T
1885/0 ALIGNMENT IS ARBITRARY
1886/0 ALIGNMENT IS ARBITRARY

Nucleic Acids Research

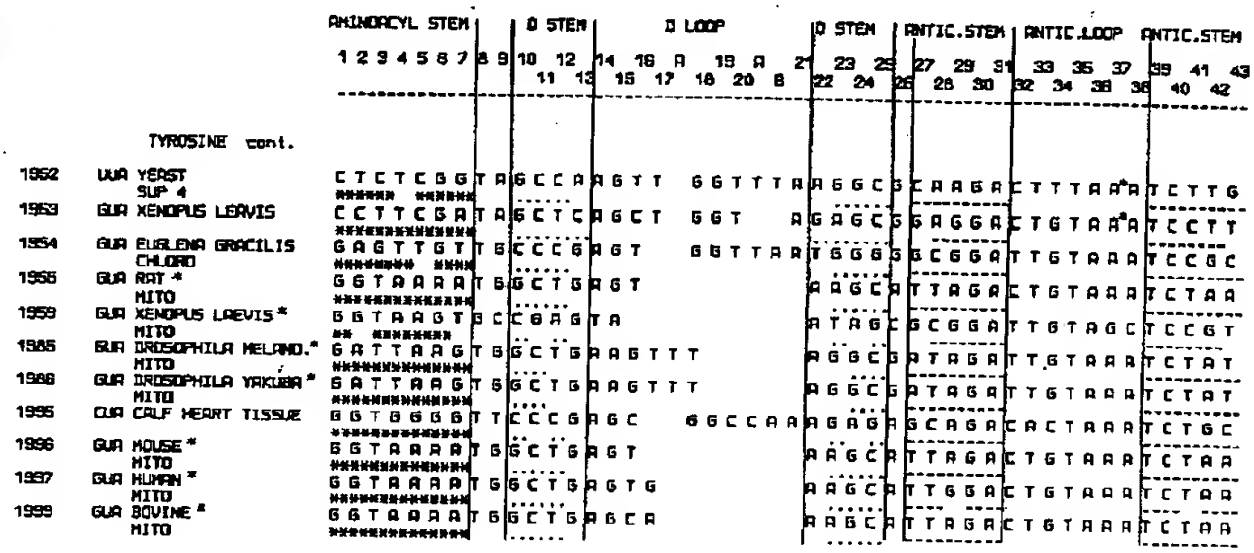


1896 B.J.BIBB, R.A.VAN ETTEN, C.T.WRIGHT, M.W.WALBERG, D.A.CLAYTON (1981) CELL 26, 167-180
1897 S.ANDERSON ET AL. (1981) NATURE 290, 457-465
1899 S.ANDERSON, A.H.L.DEBRUIJN, A.R.COULSON, I.C.EPERON, F.SANGER, I.G.YOUNG (1982) J. MOL. BIOL. 156, 683-717
1901 H.K.SCCORRLE, S.ALTMAN (1982) J. MOL. BIOL. 155, 93-103
J.J.ROSSY, A.LANDY (1979) CELL 16, 329-334
S.AN,J.D.FRIESEN(1980) GENE 12, 33-37
J.J.ROSSI,A.SCHOLD,G.F.LARSON,R.B.WALLACE(1982) GENE 20, 423-432
A.AIYAJIMA,T.YOKOTA,Y.TAKEBE,N.NAKAMURA,Y.KAZIEO(1983) J.BIOCHEM. 93, 1101-1108
1902 L.HUDSON, J.ROSSI, A.LANDY (1981) NATURE 294, 422-427
J.ROSSI,J.EGAN,L.HUDSON,A.LANDY(1983) CELL 26, 305-314
S.AN,J.D.FRIESEN(1980) GENE 12, 33-37
1905 J.J.SEILHAMER, D.J.CUMMINGS (1982) NUCL. ACIDS RES. 9, 6391-6406
1947 B.METZKER, W.G.KOECHEL, W.BAZAK, W.RUENTZEL (1982) NUCL. ACIDS RES. 10, 4763-4794
1951 E.M. DE ROBERTIS, R.V. OLSON (1979) NATURE 278, 137-143
H.R.GOODMAN, R.V.OLSON, B.D.HALL (1977) PROC. NATL. ACAD. SCI. USA 74, 5453-5457
F.F.JOHNSON,J.ABELSON(1983) NATURE 302, 681-687

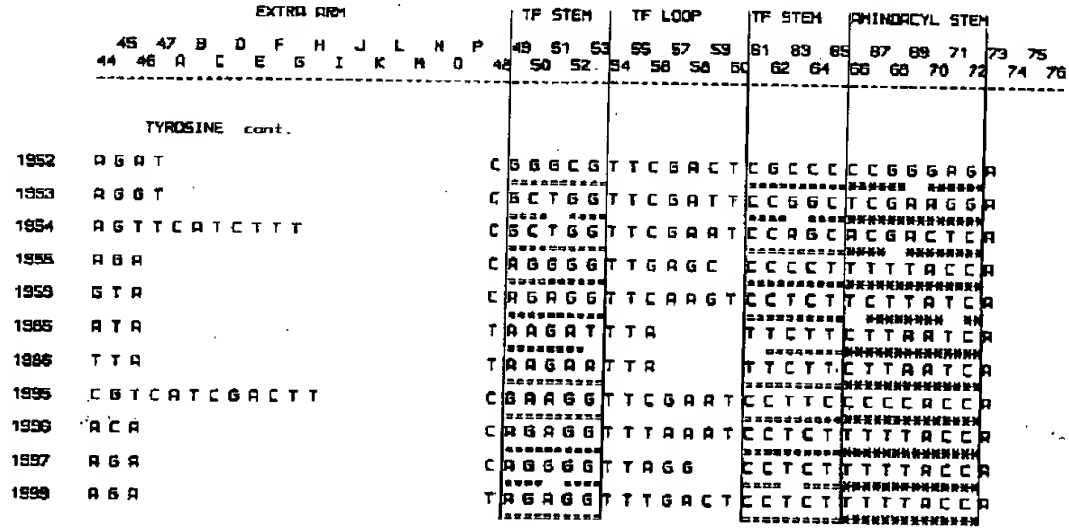


1896/0 ALIGNMENT IS ARBITRARY
1897/0 ALIGNMENT IS ARBITRARY

1899/0 ALIGNMENT IS ARBITRARY
1951/37 AFTER RESIDUE 37 INTERVENING SEQUENCE

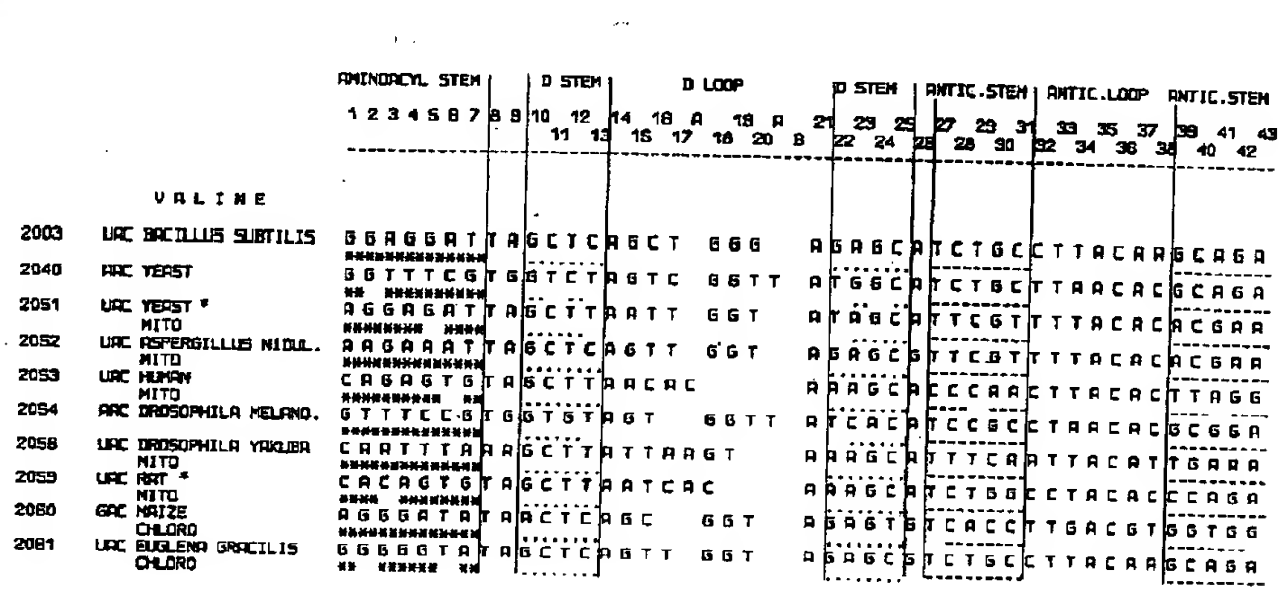


1952 D.G.ALLISON,S.H.GOH,S.O.HALL(1983) CELL 34,655-664
K.NISHIKURA,J.KUR.AN,S.O.HALL,E.A.ROBERTS(1982) EMBO J. 1,263-268
J.J.ROSSI,S.SCHOLT,G.P.LARSON,R.B.WALLACE(1982) GENE 20, 423-432
1953 F.MUELLER, B.O. CLARKSON (1980) CELL 19, 345-353
F.a.LASKI,A.Z.FIRE,U.L.RAJBHANDARY,P.A.SHARP(1983) J.BIOL.CHEM. 258, 11974-11980
1954 A.J.HOLLINGSWORTH, R.B.WALLICK (1982) J. BIOL. CHEM. 257, 2795-2799
1955 P.CANTATORE, C.DE BENEDETTO, B.GADALETA, R.BALLERANI, A.A.KADOM, M.HOLTROP, C.LANAVE, G.PEPE, C.DUAGLIARIELLO, C.SACCONE, E.SBISA (1982) NUCL. ACIDS RES. 10, 3279-3289;
R.GROSSKOPF, H.FELDMANN (1981) CURR. GEN. 4, 191-196
1959 J.F.H.WOOD, D.P.MA, R.K.WILSON, B.A.ADE (1983) NUCL. ACIDS RES. 11, 4977-5004
1985 M.H.L.DE BRUIJN(1983) NATURE 304, 234-241
1986 D.O.CLARY, D.R.WOLSTENHOLME (1983) NUCL. ACIDS RES. 11, 6859-6872
1995 E.ANDOSHI,S.ALTMANN(1983) NUCL.ACIDS RES. 11, 3687-3702
1996 R.J.BIBB, R.A.VAN ETTEN, C.T.WRIGHT, M.W.WALBERG, D.A.CLAYTON (1981) CELL 26, 167-180
1997 S.ANDERSON ET AL. (1981) NATURE 290, 457-465
1999 S.ANDERSON, M.H.L.DEBRUIJN, A.R.COULSON, I.C.EPERON, F.SANGER, I.G.YOUNG (1982) J. MOL. BIOL. 156, 683-717

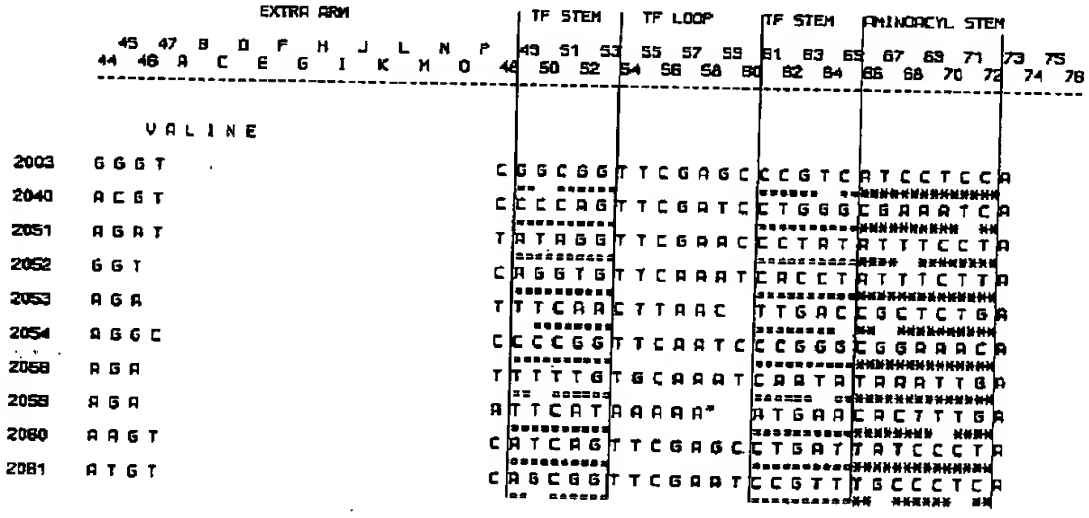


1952/37 AFTER RESIDUE 37 INTERVENING SEQUENCE
1953/37 AFTER RESIDUE 37 INTERVENING SEQUENCE
1955/0 ALIGNMENT IS ARBITRARY
1959/0 ALIGNMENT IS ARBITRARY
1985/0 ALIGNMENT IS ARBITRARY
1986/0 ALIGNMENT IS ARBITRARY
1996/0 ALIGNMENT IS ARBITRARY
1997/0 ALIGNMENT IS ARBITRARY
1999/0 ALIGNMENT IS ARBITRARY

Nucleic Acids Research

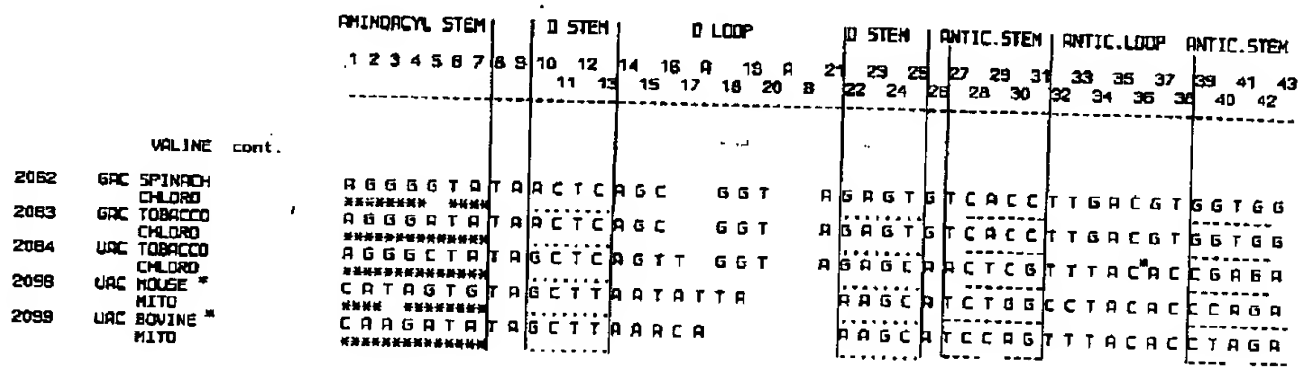


2003 C.J.GREEN, B.S.VOLD (1983) NUCL. ACIDS RES. 11, 5763-5774
2040 R.E.BAKER, A.EIGEL, O.VOEGTEL, H.FELDMANN (1982) EMBO JOURNAL 1, 291-295
2051 O.MILLER, C.SIGURDSON, N.C.MARTIN, J.E.DOWELSON (1980) NUCL. ACIDS RES. 8, 1425-1442
M.LI, A.TZAGOLOFF (1979) CELL 18, 47-53
2052 H.D.KOECHEL, C.R.LAZARUS, N.BASAK, H.KUENTZEL (1981) CELL 23, 625-633
2053 I.C.EPERON, S.ANDERSON, D.P.NIERLICH (1980) NATURE 286, 460-467
ANDERSON ET AL. (1981) NATURE 290, 457-465
2054 W.R.ADDISON, C.R.ASTELL, A.D.DELANEY, I.A.GILLAN, S.HAYASHI, W.C.MILLER, B.RAJPUT, R.SMITH, D.R.TAYLOR, D.R.TENER (1982) J. BIOL. CHEM. 257, 670-673
2058 D.O.CLARY, J.B.GODDARD, S.G.MARTIN, E.R.-R.FAURON, D.R.WOLSTENHOLME (1982) NUCL. ACIDS RES. 10, 6619-6637
2059 C.SACCONE, F.CANTATORE, G.GADALETA, R.GALLERANI, C.LANAVE, G.PEPE A.R.KRSON (1981) NUCL. ACIDS RES. 9, 4131-4148
S.KOBAYASHI, T.SEKI, K.YAGINUMA, K.KOIKE (1981) GENE 14, 297-307
2060 Z.SCHWARZ, H.KOESSEL, C.SCHWARZ, L.BOGORAD (1982) PROC. NATL. ACAD. SCI. USA 78, 4748-4752
2081 E.R.OROZCO, JR, I.G.HALLICK (1982) J. BIOL. CHEM. 257, 3265-3275

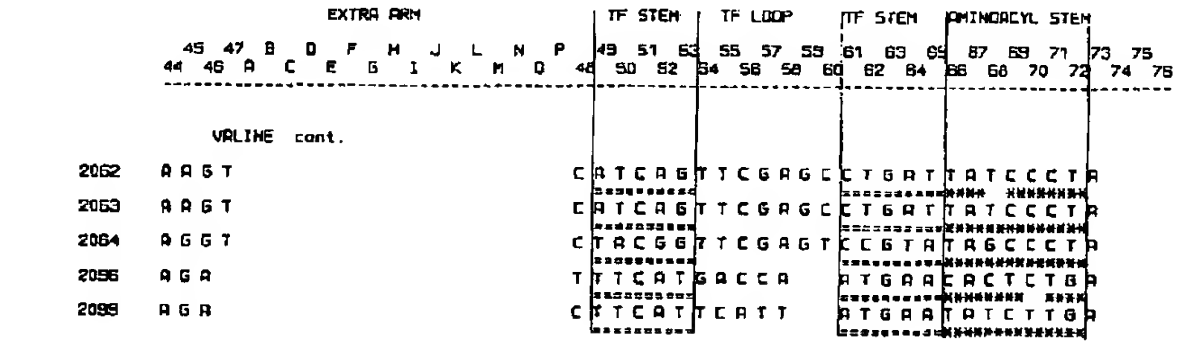


2051/0 IN B.E.TRAMLENFELD,A. IZAGOLOFF (1980) J.BIOL.CHEM. 255, 6173-6180
SEQUENCE WITH DEVIATING 39-42

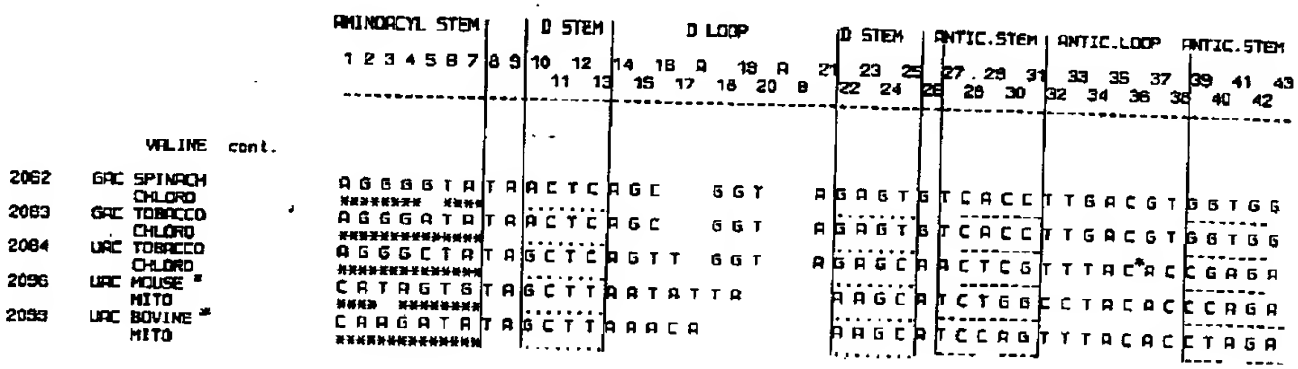
2037/0 ALIGNMENT IS ARBITRARY
2059/58 S.KOBAYASHI ET.AL.DID NOT FIND A AT THIS POSITION



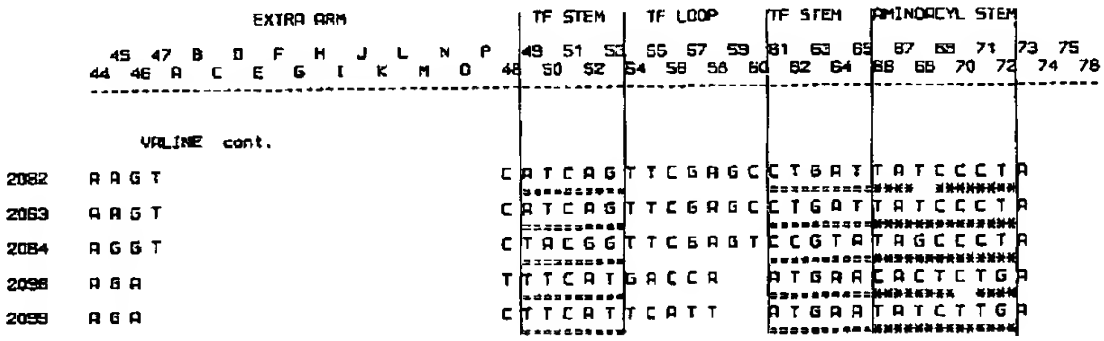
2062 J.-F.BRIAT,A.DHDH,S.LOISEAUX,R.GACHE(1982) NUCL.ACIDS RES. 10, 5365-5378
2063 M.TOHDOH, K.SHINOZAKI, A.SUGIURA (1981) NUCL. ACIDS RES. 9, 5399-5406
2064 H.DEND, A.KATO, K.SHINOZAKI, A.SUGIURA (1982) NUCL. ACIDS RES. 10, 7511-7520
2096 A.J.BIBB, R.A.VAN ETTEN, C.T.WRIGHT, M.D.WALBERG, D.A.CLAYTON (1981) CELL 26, 167-180
2099 S.ANDERSON, A.N.L.DEBRUIJN, A.R.COULSON, I.C.EPERON, F.SANGER, I.G.YOUNG (1982) J. MOL. BIOL. 156, 683-717



2064/36 AFTER RESIDUE 36 OR 37 INTERVENING SEQUENCE
2096/0 ALIGNMENT IS ARBITRARY
2099/0 ALIGNMENT IS ARBITRARY



2062 J.-F.BRIAT,M.GRON,S.LOISEAUX,A.MACHE(1982) NUCL.ACIDS RES. 10, 6865-6875
2063 N.TORDOH, K.SHINOZAKI, M.SUGIURA (1981) NUCL. ACIDS RES. 9, 5399-5404
2064 H.DEND, A.KATO, K.SHINOZAKI, M.SUGIURA (1982) NUCL. ACIDS RES. 10, 7511-7520
2096 R.J.BIBB, R.A.VAN ETTEN, C.T.WRIGHT, M.W.WALBERG, D.A.CLAYTON (1981) CELL 26, 167-180
2099 S.ANDERSON, A.R.L.DEBRUIJN, A.R.COULSON, I.C.EPERON, F.SANGER, I.G.YOUNG (1982) J. MOL. BIOL. 156, 683-717



2064/36 AFTER RESIDUE 34 OR 37 INTERVENING SEQUENCE
2096/0 ALIGNMENT IS ARBITRARY
2099/0 ALIGNMENT IS ARBITRARY

Compilation of tRNA sequences and sequences of tRNA genes

Mathias Sprinzl*[1], Konstantin S. Vassilenko[2]

[1] Laboratorium für Biochemie, Universität Bayreuth, 95440 Bayreuth, Germany and [2] Institute of Protein Research, Russian Academy of Sciences, 142290 Puschchino, Moscow Region, Russia

*To whom correspondence should be addressed

Tel.: +49 921 552420

Fax: +49 921 552432

email:

mathias.sprinzl@uni-bayreuth.de;

kvassil@vega.protres.ru

INTRODUCTION

The new compilation of tRNA Sequences and Sequences of tRNA genes contains in addition to 3279 sequences of the last edition from 1998 (1) the completely new Genomic tRNA Compilation including the sequences of tRNA genes from complete genomes published up to January 2002. The current Database consists of three parts:

1. Compilation of tRNA Genes (MS Excel® file, ZIPed)

2. Compilation of tRNA Sequences (MS Excel® file, ZIPed)

3. Genomic tRNA Compilation (MS Excel® file, ZIPed)

Compilation of tRNA Genes,

is a summary of the sequences of tRNA genes published in the literature and databases up to the end of 1998. It contains tRNA genes of all organisms and organels, but is not updated since January 1999. This table contains about 500 sequences of cytoplasmic tRNA genes that are not included in the Genomic tRNA Database. Most of the tRNA gene entries in this table have references of the publications in which the sequence was communicated.

Compilation of tRNA Sequences,

is a summary of tRNA sequences, including modified bases and references of the publications. The references are restricted to the first publication of the complete sequence unless additional information (e.g. base modification, corrections, etc.) was later obtained. In such cases additional references were added. This compilation is updated up to January 2002. The table contains the known tRNA sequences of all organisms including organells. This is the continuation of the original tRNA compilation first published in 1978.

Genomic tRNA Compilation,

is a new addition to the Database. This is the most complete compilation of the sequences of cytoplasmic tRNA genes derived from complete genome sequences included into DNA databases. Since sequences of tRNA genes originated from cellular organelles frequently can not be processed to the general cloverleaf scheme, they were not included in the Genomic tRNA Compilation. There are specialised databases dealing with these sequences (see links below).

Current Genomic tRNA Compilation consists of about 3700 tRNA gene sequences from 63 organisms covering archaea, bacteria, higher and lower eukarya. The database includes the tRNA genes sequences collected in GtRDB (2) as well as those from the additional complete genomes found in DNA databases. If the genomes of the different strains of the same organism were sequenced, the corresponding tRNA genes were added to the database independently.

PRESENTATION OF SEQUENCES

Compilation of tRNA Genes and Compilation of tRNA Sequences

In order to facilitate a computer analysis an alignment of sequences is used which is most compatible with the tRNA phylogeny and known three-dimensional structures of tRNA (3, 4). The corresponding numbering system is shown in Figure 1. Positions in particular sequence which are not filled (gaps in the generalised structure) are indicated by a dash. All nucleotide insertions are commented and denoted by underlining at the place of insertion.

This compilations use a one-letter code for all nucleotides including modified ones. For standard nucleotides, adenosine, cytidine, guanosine, thymidine and uridine the usual abbreviations, A, C, G, T and U, respectively, are used. To designate modified nucleotides, the other ASCII signs are employed (see table "Intro" in the corresponding MS Excel® file). Terminology and structure of the modified nucleosides occurring in tRNAs were used according to (5) and (6).

Sequences are presented as MS Excel® files. Each sequence in the compilation occupies two consecutive rows. The first row begins with the unique six-position identification code of the sequence ('D' or 'R' for DNA or RNA, respectively; a one-letter code for the amino acid, X for methionine-initiator, Z for selenocysteine; the three-digit code specifying the organism and one digit for isoacceptor number). After this, the sequence of the anticodon is given, followed by the abbreviated name and the kingdom of organism, and the sequence (99 standard positions). The second line begins with the sign '+' and contains the information about base-pairing (double helical regions only, tertiary interactions are not annotated). Nucleotides involved in Watson-Crick pairs are marked with '=', the GU pairs are indicated with the sign '*'.

Genomic tRNA Compilation

The database is organised as an MS Excel® workbook. All the information collected are split into different indexed tables according to the type of data (specificity, sequence, organism, etc.) and the descriptions of certain genes are summarised in the main worksheet that includes the relations between the data tables. The information can be obtained by filling the query form that allows to enter the simple search criteria and to select the type of data to be displayed. The result of search is presented as a table containing the description of the genes found. This includes unique id, amino acid specificity, anticodon sequence, organism name and taxonomy, strain, original database source, position of the gene in genome, literature reference, sequence, basepairing and additional comments. Sequences are aligned in the same way as it was described above for the tRNA compilations.

In addition to the plain text table one can explore the result of search by presenting the sequences in a cloverleaf form (Figure 1). It is possible to scroll the found sequences one by one or to select directly the sequence of interest from the result table. The presentation supports colour code for different structural features in the canonical cloverleaf model.

Simple statistical information on the occurrences of certain bases at given positions and the preferences in basepairing also can be obtained on a special data sheet.

Useful links:

The RNA Modification Database
http://medlib.med.utah.edu/RNAmods

A database for plant mitochondrial tRNA genes and molecules
http://bio-www.ba.cnr.it:8000/BioWWW/#PLMItRNA

Compilation of mammaliam mitochondrial tRNA genes
http://mamit-trna.u-strasbg.fr

GtRDB: The Genomic tRNA Database
http://rna.wustl.edu/tRNAdb

ACKNOWLEDGEMENT

This project was supported by Fonds der Chemischen Industrie and Universität Bayreuth. We are gratefull for advise, cooperation and help with data collection to Todd Michael Johnson Lowe, Genetics, Stanford University, California, Dr. Carlos Hoyo-Vadillo, Departamento de Farmacologia y Toxicologia, Cinvestav, Mexico City, Mexico, Yvonne Baberowski, Mark Dürr and Bernhard Thielen, Universität Bayreuth.

REFERENCES

1. Sprinzl M., Horn C., Brown M., Ioudovitch A. and Steinberg S. (1998) *Nucl. Acids Res.* 26, 148-153.

2. Lowe, T.M. & Eddy, S.R. (1997) *Nucl. Acids Res.* 25, 955-964.

3. Transfer-RNA: Structure, properties and recognition, P.R. Schimmmel, D. Söll, J.N. Abelson Eds., 1979, Cold Spring Harbor Laboratory, p.518-519.

4. Steinberg S.V. and Kisselev L.L. (1992) *Biochimie* 74, 337-351.

5. Limbach P.A., Crain P.F. and McCloskey, J.A. (1994) *Nucl. Acids Res.* 22, 2183-2196.

6. Crain P.F. and McCloskey J.A. (1997) *Nucl. Acids Res.* 25, 126-127.

tRNA database searching engine

Internet service that allows to find records in the database according to multiple search criteria. Complicated sequence-based queries can be formed (Updated for the data in Compilation of tRNA Genes and Compilation of tRNA Sequences up to the end of 1998).

tRNA-Editor

Researchers who wish to perform an advanced search for tRNA sequences according to several criteria, e.g. anticodon, amino acid specificity, modified nucleoside, or wish to print the requested sequences in the cloverleaf form can download appropriate Windows 3.1 based software as a 900kB ZIPed file (Updated for the data in Compilation of tRNA Genes and Compilation of tRNA Sequences up to the end of 1998).

EXHIBIT F

A Rare tRNA-Arg(CCU) That Regulates Ty*1* Element Ribosomal Frameshifting Is Essential for Ty*1* Retrotransposition in *Saccharomyces cerevisiae*

Koichi Kawakami,* Suchira Pande,† Brenda Faiola,‡ Daniel P. Moore,§ Jef D. Boeke,§ Philip J. Farabaugh,† Jeffrey N. Strathern,‡ Yoshikazu Nakamura* and David J. Garfinkel‡

**Department of Tumor Biology, The Institute of Medical Science, University of Tokyo, Tokyo 108, Japan, †Department of Biological Sciences, University of Maryland-Baltimore County, Catonsville, Maryland 21228, ‡Laboratory of Eukaryotic Gene Expression, NCI-Frederick Cancer Research and Development Center, ABL-Basic Research Program, Frederick, Maryland 21701-1201, and §Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205*

Manuscript received February 24, 1993
Accepted for publication June 10, 1993

ABSTRACT

Translation of the yeast retrotransposon Ty*1* *TYA1(gag)-TYB1(pol)* gene occurs by a +1 ribosomal frameshifting event at the sequence **CUU AGG C**. Because overexpression of a low abundance tRNA-Arg(CCU) encoded by the *HSX1* gene resulted in a reduction in Ty*1* frameshifting, it was suggested that a translational pause at the AGG-Arg codon is required for optimum frameshifting. The present work shows that the absence of tRNA-Arg(CCU) affects Ty*1* transposition, translational frameshifting, and accumulation of mature TYB1 proteins. Transposition of genetically tagged Ty*1* elements decreases at least 50-fold and translational frameshifting increases 3–17-fold in cells lacking tRNA-Arg(CCU). Accumulation of Ty*1*-integrase and Ty*1*-reverse transcriptase/ribonuclease H is defective in an *hsx1* mutant. The defect in Ty*1* transposition is complemented by the wild-type *HSX1* gene or a mutant tRNA-Arg(UCU) gene containing a C for T substitution in the first position of the anticodon. Overexpression of *TYA1* stimulates Ty*1* transposition 50-fold above wild-type levels when the level of transposition is compared in isogenic *hsx1* and *HSX1* strains. Thus, the *HSX1* gene determines the ratio of the TYA1 to TYA1-TYB1 precursors required for protein processing or stability, and keeps expression of *TYB1* a rate-limiting step in the retrotransposition cycle.

THE *Saccharomyces cerevisiae* retrotransposon Ty*1* is a mobile genetic element that replicates via an RNA intermediate (reviewed by BOEKE and SANDMEYER 1991; GARFINKEL 1992). The transposition cycle of Ty*1* elements resembles several important steps in the replication of retroviruses. Ty*1* protein maturation by Ty*1*-protease (PR) and reverse transcription take place within Ty*1* virus-like particles (Ty*1*-VLPs), which appear to be absolutely required for the transposition process. The Ty*1* genome contains two genes, *TYA1* and *TYB1*, which correspond to the *gag* and *pol* genes of retroviruses, respectively (CLARE and FARABAUGH 1985). As with certain retroviral *pol* genes (reviewed by HATFIELD *et al.* 1992), expression of *TYB1* requires programmed ribosomal frameshifting (CLARE, BELCOURT and FARABAUGH 1988). Ribosomal frameshifting solves two problems encountered in the life cycle of a retrovirus or retrotransposon. First, since catalytic Pol proteins, such as reverse transcriptase/ribonuclease H (RT/RH) and integrase (IN), are usually found in much lower amounts than the structural Gag proteins, requiring a frameshift event for *pol* expression is an effective strategy of gene regulation. Second, since Pol proteins function within a particle, creating a Gag-Pol fusion protein by frameshifting delivers Pol proteins to the correct compartment.

The TYA1-TYB1 fusion protein is synthesized by a +1 frameshifting event in the TYA1 sequence **CUU AGG C** (BELCOURT and FARABAUGH 1990). Ribosomal pausing at a rare AGG-arginine codon and slippage of a leucyl-tRNA from CUU to UUA are required for frameshifting. A single-copy tRNA-Arg(CCU) gene that recognizes the AGG codon is located on chromosome *X* (GAFNER, DE ROBERTIS and PHILIPPSEN 1983). BELCOURT and FARABAUGH (1990) have shown that overexpression of the tRNA-Arg(CCU) gene reduces Ty*1* frameshifting. Ty*1* transposition is also reduced when the level of the tRNA-Arg(CCU) is increased (XU and BOEKE 1990). These results suggest that the low abundance of tRNA-Arg(CCU) promotes frameshifting. Recently, we have identified this tRNA gene as the *HSX1* gene involved in the heat shock response (KAWAKAMI *et al.* 1992). Even though there is only one copy of the *HSX1* gene (GAFNER, DE ROBERTIS and PHILIPPSEN 1983), an *hsx1* disruption mutant is viable. Apparently, the AGG codons normally decoded by the single-copy *HSX1* gene are decoded by another tRNA [probably by the near-cognate tRNA-Arg(UCU)

TABLE 1

Yeast strains

Strain	Genotype	Plasmid	Source or reference
DMy51	*MATα ura3-167 his3Δ200 leu2Δ trp1Δ1 GAL*	pGTy*1*A-B*neo* (pD109)	This work
DMy94	*MAT***a** *ura3-52 his3Δ200 lys2 trp1-289 GAL*		This work
JC287	*Matα ura3-167 his3Δ200 leu2GB* Ty*1*-146::*lacZ* Ty*1*m*his3*AI-263 *GAL*		M. J. CURCIO
JC344	*MATα ura3-167 his3Δ200 leu2GB* Ty*1*-146::*lacZ* Ty*1*m*his3*AI-270 *GAL*		M. J. CURCIO
KK156	JC287; *hsx1::LEU2*		This work
KK157	JC344; *hsx1::LEU2*		This work
KK240	*MAT***a** *ura3 his3 leu2 trp1 hsx1::HIS3*		This work
KK242	*MAT***a** *ura3 his3 leu2 trp1*		This work
KD198-16A	*MAT***a** *his4Δ5 ura3 arg11 GAL*		K. J. DURBIN
DG1301	JC344	p*GAL1-lacZ*	This work
DG1302	JC344	pGTy*1*-H3*neo*	This work
DG1305	KK157	p*GAL1-lacZ*	This work
DG1306	KK157	pGTy*1*-H3*neo*	This work
DG1333	JC344	pGTy*1*-H3*neo*::*SacI-1702*	This work
DG1334	KK157	pGTy*1*-H3*neo*::*SacI-1702*	This work
DG1344	JC344	pGTyA1*neo*(*PGK1* ter.)	This work
DG1347	KK157	pGTyA1*neo*(*PGK1* ter.)	This work

gene]. In this paper, we describe the effects of an *hsx1* disruption mutant on Ty*1* frameshifting, transposition and protein processing.

MATERIALS AND METHODS

Yeast strains, plasmids, general genetic methods and media: The strains used for the Ty*1* transposition assays are listed in Table 1. Strains KK240 (*MAT***a** *ura3 his3 leu2 trp1 hsx1::HIS3*) and KK242 (*MAT***a** *ura3 his3 leu2 trp1*) were used to test Ty*1* frameshifting. These strains were derived from an *hsx1::HIS3/HSX1* diploid strain (KAWAKAMI *et al.* 1992).

The plasmids pMB38-9merWT and pMB38-9merFusion contain the frameshift heptamer fused to *Escherichia coli lacZ* gene in the +1 *TYB1* reading frame and the 0 *TYA1* reading frame, respectively (BELCOURT and FARABAUGH 1990). The plasmid pMB38-9merFusion(w/o AGG) contains the AGG-less 0 reading frame [GAT CCG CTG ACA **CTT GGC** CAT GAG GTA C (the frameshift region is highlighted)] fused to *lacZ*. The plasmid pKK67 was constructed by cloning the 230 base-pair (bp) wild-type *HSX1* DNA, amplified by polymerase chain reaction (PCR) (SAIKI *et al.* 1985) into the *URA3*-based centromere-plasmid YCp50 (ROSE, *et al.* 1987). The plasmid pKK68 carrying the mutant *hsx1*(*Mlu*I*) gene was constructed by digestion of the plasmid pKK67 with *Mlu*I, fill-in synthesis with Klenow DNA polymerase, and ligation to a *Sal*I linker. The *hsx1::HIS3* and *hsx1::LEU2* disruption alleles were constructed by modifying the same *Mlu*I restriction site and ligation to a *Cla*I fragment containing the *HIS3* gene, or an *Mlu*I-*Cla*I fragment containing the *LEU2* gene (kindly provided by P. ROGAN). The plasmid pKK69 was constructed by cloning the PCR-amplified 112-bp wild-type *SUP201-0* gene (THIREOS, PENN and GREER 1984; MORISHITA and UNO 1991) into the *URA3*-based centromere-plasmid pRS316 (SIKORSKI and HIETER 1989). The plasmid pKK71 carrying the *SUP201-0-1*(*CCU*) gene was constructed by digestion of plasmid pKK69 with *Mlu*I and *Bam*HI and ligation to a 63-bp synthetic double-stranded DNA containing the C for T substitution at 3′ base of the anticodon. The *Eco*RI-*Bam*HI DNA fragments containing the mutant and wild-type tRNA genes were prepared from plasmids pKK67, pKK68, pKK69 and pKK71, and subcloned into the *TRP1*-based centromere-plasmid pRS314 (SIKORSKI and HIETER 1989). These subcloning procedures generated plasmids pKK73 (derived from plasmid pKK67), pKK74 (from pKK68), pKK75 (from pKK69), and pKK76 (from pKK71). The plasmid pGTy*1*A-B*neo* (also known as plasmid pD109), with the Ty*1* frameshift correctly removed (BELCOURT and FARABAUGH 1990), was constructed from a transposition-competent pGTy*1*-H3/Ty*1*-912 hybrid plasmid by oligonucleotide-bridge mutagenesis (MANDECKI 1986). The frameshift mutation and tRNA sequences were confirmed by chain-terminating DNA sequencing (SANGER, NICKLEN and COULSON 1977) using Sequenase 2.0 (U.S. Biochemical Corp.). The plasmid pGTyA*1neo* (PGK ter.), kindly provided by P. ROGAN, was constructed by replacing almost all of the pGTy*1*-H3 TYB1 gene (from a *Bgl*II site located at position 1702 to the end of the element) with the bacterial neo gene and the *PGK1* transcriptional terminator. Standard techniques were used for all molecular cloning procedures (SAMBROOK, FRITSCH and MANIATIS 1989).

The *hsx1::HIS3* and *hsx1::LEU2* disruption mutants were constructed by single-step gene disruption (ROTHSTEIN 1991). Plasmids were introduced into cells using the transformation procedure of ITO *et al.* (1983). All yeast media and standard genetic techniques were those described by ROSE, WINSTON and HIETER (1990).

Transposition assays: Ty*1*m*his3*AI and Ty*1*m*ade2*AI transposition assays were performed as described previously (CURCIO and GARFINKEL 1991, 1992), and will be presented briefly here. For detecting spontaneous Ty*1*m*his3*AI transposition events, liquid cultures were inoculated at low densities (about 2×10^5 cells/ml) and grown to saturation at 20° in YPD or in SC-ura (glucose). A portion of each culture was spread on SC-his or SC-his-ura (glucose) plates and incubated at 30°. The cultures were titered on YPD or SC-ura (glucose) plates. For detecting chromosomal Ty*1*m*his3*AI transposition events in the presence of a pGTy*1* helper plasmid, cells were grown on SC-ura (galactose) plates for 7 days at 20°, or an overnight SC-ura (glucose) liquid culture was diluted 50-fold into SC-ura (galactose) liquid medium and incubated with aeration for

3 days at 20°. Ty*1*m*his3*AI transposition events were detected as His$^+$ papillae by replica plating cells from the SC-ura (galactose) to SC-his-ura (glucose) plates, followed by incubation at 30° for 3 days. To determine the number of Ty*1*m*his3*AI or Ty*1*m*ade2*AI transposition events in galactose-grown liquid cultures, the cells were concentrated, spread on several SC-his-ura (glucose) or SC-ade-ura (glucose) plates, and incubated at 30° for 3–5 days. Cells were titered on SC-ura (glucose) plates. Ty*1*n*eo* and Ty*1*A-B*neo* transposition events were detected as described previously (BOEKE, XU and FINK 1988; CURCIO, SANDERS and GARFINKEL 1988) with the following minor modifications. Diploid strains were constructed by mating strains KD198-16A with strains DG1302 or DG1306, or by mating strains DMy51 and DMy92 (Table 1). The resulting diploids were induced for Ty*1* transposition on SC-ura (galactose) plates as described above. After segregation of the pGTy*1neo* plasmid from the strains, the level of resistance to the antibiotic G418 (Gibco) was determined by growth on YPD plates containing a final G418 concentration of 500 μg per ml (for diploids derived from mating strain KD198-16A with strains G1302 or DG1306) or 75 μg per ml (for diploid derived from mating strain DMy51 with DMy94).

Ty*1* RNA levels and Ty*1*m*his3*AI splicing efficiency: We isolated total RNA from *hsx1* and *hsx1::LEU2* strains by established procedures (CURCIO, SANDERS and GARFINKEL 1988; ROSE, WINSTON and HIETER 1990). Northern analysis was used to analyze Ty*1* RNA levels (CURCIO, SANDERS and GARFINKEL 1988; CURCIO and GARFINKEL 1992), and reverse transcription-PCR (RT-PCR) was used to estimate Ty*1*m*his3*AI RNA splicing efficiency (WANG, DOYLE and MARK 1989). The total amount of RNA transferred to hybridization membranes was estimated by staining with NAQ-STAIN, a reversible fluorescein-based stain developed by Integration Separation Systems. Transcripts from the *PYK1* and *ACT1* genes were used as internal loading standards. RNA sequences that span the region where the artificial intron (AI) was inserted in *HIS3* (CURCIO and GARFINKEL 1991) were amplified using the *HIS3*-specific oligonucleotide primers CTCCACGCGCCAGTAGGGCC (for DNA amplification) and ATGACAGAGCAGAAAGC CC (for reverse transcription and DNA amplification). The amplified products were separated by agarose gel electrophoresis through a 2% NuSieve/1% SeaKem (FMC Bioproducts) gel, stained with ethidium bromide, and photographed. The resulting negatives were scanned using an LKB Ultroscan XL enhanced laser densitometer. Relative splicing efficiencies were estimated by the amount of the amplified products. The splicing efficiency is defined as the amount of 334-bp spliced product over the amount of spliced plus 438-bp unspliced products.

Immunoblot analysis: Total yeast protein isolation, polyacrylamide gel electrophoresis, protein transfer, and antibody reactions were performed as described previously (YOUNGREN *et al.* 1988; GARFINKEL *et al.* 1991). Antibodies were added in at least 10-fold excess, as determined by titration experiments. Ty*1*-VLP antibodies were previously shown to react with TYA1 and TYA1-TYB1 precursor proteins, but not with TYB1 proteins (ADAMS *et al.* 1987; YOUNGREN *et al.* 1988; GARFINKEL *et al.* 1991). Ty*1*-VLP antibodies did not show a dramatic difference in avidity for TYA1 *vs.* TYA1-TYB1 precursor proteins, as determined by titration experiments (A.-M. HEDGE and D. J. GARFINKEL, unpublished results). Ty*1*-VLPs were isolated by the method of EICHINGER and BOEKE (1988), except the final continuous sucrose gradient was omitted. Equal amounts of protein (approximately 20 μg per lane) were loaded onto SDS-8% polyacrylamide gels. Protein concentrations were verified by staining gels run in parallel with Coomassie blue. Cross-reactivity of immunoblotted proteins with antisera that recognize the mature proteins p54-TYA1 (Ty*1*-VLP antiserum; ADAMS *et al.* 1987; YOUNGREN *et al.* 1988), p90-Ty*1*-IN (B2 antiserum; YOUNGREN *et al.* 1988), p60-Ty*1*-RT/RH (B8 antiserum; GARFINKEL *et al.* 1991), and their respective precursor proteins were detected using the ECL chemiluminescent detection system (Amersham).

Ty*1* frameshifting efficiency: β-Galactosidase assays and the efficiency of Ty*1* frameshifting were determined as described previously (BELCOURT and FARABAUGH 1990). Briefly, six transformants of each plasmid were each assayed in triplicate for β-galactosidase activity. The frameshifting efficiency is measured by determining the ratio of β-galactosidase activity produced from the construct requiring a +1 frameshift to express *lacZ* (pMB38-9merWT) to that of a construct in which the upstream and downstream genes are fused in frame [pMB38-9merFusion and pMB38-9merFusion(w/oAGG)].

The efficiency of Ty*1*-H3 frameshifting was also estimated from immunoblot analysis. Strains DG1333 (pGTy*1*-H3*neo::SacI-1702, HSX1*) and DG1334 (pGTy*1*-H3*neo::SacI-1702, hsx1::LEU2*) were constructed by transforming the plasmid pGTy*1*-H3*neo::SacI-1702*, which contains a Ty*1*-PR mutation (YOUNGREN *et al.* 1988), into strains JC344 and KK157, respectively (Table 1). Total protein isolated from galactose-grown cultures of strains DG1333 and DG1334 was analyzed by immunoblotting using Ty*1*-VLP antiserum. To determine the ratio of p58-TYA1 to p190-TYA1-TYB1 protein, exposures of the resulting blots were scanned using a laser densitometer. The efficiency of Ty*1* frameshifting equals the amount of p190-TYA1-TYB1 protein divided by the total amount of p58-TYA1 plus p190-TYA1-TYB1 protein.

RESULTS

Ty*1* transposition is inhibited in an *hsx1* disruption mutant: We determined whether a disruption mutation of *HSX1* affects Ty*1* transposition using two assays that monitor transposition of chromosomal elements marked with the *his3*AI retrotransposition indicator gene (CURCIO and GARFINKEL 1991), as well as by monitoring the transposition of plasmid-borne pGTy*1neo* and pGTy*1made2*AI elements (BOEKE *et al.* 1985; BOEKE, XU and FINK 1988; M. J. CURCIO and D. J. GARFINKEL, unpublished results). The *his3*AI gene is a yeast *HIS3* gene interrupted by an artificial intron (AI) in the antisense orientation. The *his3*AI sequences are inserted in a Ty*1* element at a unique restriction site located between the *TYB1* gene and the downstream long terminal repeat, such that the intron is on the sense strand of the Ty*1* element. Placement of marker genes at this position of a Ty*1* element does not severely inhibit transposition. Since splicing and retrotransposition of the marked Ty RNA gives rise to His$^+$ cells, the relative efficiency of Ty*1*m*his3*AI transposition can be monitored by plating cells on media lacking histidine. An *ade2*AI retrotransposition indicator gene has also been developed (M. J. CURCIO and D. J. GARFINKEL, unpublished results).

First, the relative efficiency of Ty*1*m*his3*AI trans-

TABLE 2

Ty*1*m*his3*AI transposition in an *hsx1* disruption mutant

Genotype	Ty*1*m*his3*AI	His$^+$ colonies/ total cells ($\times 10^7$)	Relative transposition efficiency
HSX1	Ty*1*m*his3*AI-263	25/1.6	2.0×10^{-6}
		46/1.6	
		40/1.8	
		36/1.9	
		30/1.9	
hsx1::LEU2	Ty*1*m*his3*AI-263	0/2.3	2.7×10^{-8}
		0/2.3	
		1/2.2	
		2/2.2	
		0/2.1	
HSX1	Ty*1*m*his3*AI-270	28/1.6	2.0×10^{-6}
		34/1.6	
		22/1.4	
		32/1.4	
		36/1.6	
hsx1::LEU2	Ty*1*m*his3*AI-270	3/1.7	3.8×10^{-8}
		0/2.3	
		0/2.4	
		0/2.1	
		1/1.8	

The Ty*1*m*his3*AI-263 element is present in *HSX1* strain JC287 and *hsx1::LEU2* strain KK156. The Ty*1*m*his3*AI-270 element is present in *HSX1* strain JC344 and *hsx1::LEU2* strain KK157. Each measurement represents the results of one of five independent cultures. The relative transposition efficiency is the mean fraction of total colonies that are His$^+$. To estimate the efficiency of Ty*1* transposition, the relative transposition efficiency should be multiplied by a factor of 8, to account for the splicing efficiency of the Ty*1*m*his3*AI transcript, and by a factor of 11, to account for the effect of introducing the *his3*AI marker gene into a Ty*1* element (CURCIO and GARFINKEL 1991).

TABLE 3

Ty*1*m*his3*AI-270 transposition in *hsx1* mutant KK157 containing plasmid copies of tRNA genes

Plasmid (genotype)	His$^+$ colonies/ total colonies ($\times 10^6$)	Relative transposition efficiency
pKK67 (*HSX1*)	17/5.3	6.4×10^{-6}
	58/2.3	
	16/3.6	
	18/5.3	
	23/4.1	
pKK68 [*hsx1*(*Mlu1**)]	0/4.5	$<4.9 \times 10^{-8}$
	0/4.6	
	0/4.4	
	0/2.3	
	0/4.5	
pKK69 [*SUP201-0*(*UCU*)]	0/5.7	$<3.5 \times 10^{-8}$
	0/6.5	
	0/5.2	
	0/5.6	
	0/5.9	
pKK71 [*SUP201-0-1*(*CCU*)]	4/7.5	1.1×10^{-6}
	3/5.6	
	7/5.3	
	11/5.9	
	8/6.1	

The Ty*1*m*his3*AI-270 element is present in the *hsx1::LEU2* strain KK157. The designated plasmids were introduced into strain KK157 and single transformants were chosen for further analysis. Refer to Table 2 for more information.

position in isogenic *HSX1* and *hsx1::LEU2* strains containing single marked chromosomal elements Ty*1*m*his3*AI-263 or Ty*1*m*his3*AI-270 was determined (Table 2). These unspliced Ty*1*m*his3*AI elements were identified after galactose-induction of a strain containing plasmid pGTy*1*-H3m*his3*AI, and are present at different chromosomal locations (CURCIO and GARFINKEL 1991). There was a 53- or 74-fold decrease in the efficiency of Ty*1*m*his3*AI-263 or Ty*1*m*his3*AI-270 transposition, respectively, as monitored by the number of His$^+$ colonies in a *hsx1::LEU2* mutant background. The transposition defect in the *hsx1::LEU2* mutant KK157 was complemented by a low copy number plasmid carrying the wild-type *HSX1* gene (pKK67), but not by a plasmid carrying a mutant *hsx1*(*M1u1**) gene (pKK68) (Table 3).

The second transposition assay depends upon the ability of a pGTy*1* helper plasmid to stimulate transposition of a genomic Ty*1*m*his3*AI element in trans (CURCIO and GARFINKEL 1992). Expression of the pGTy*1*-H3 helper plasmid increases the frequency of genomic Ty*1*m*his3*AI transposition about 100-fold (CURCIO and GARFINKEL 1992; M. J. CURCIO and D. J. GARFINKEL, unpublished results). The pGTy*1*-H3*neo* helper plasmid (BOEKE, XU and FINK 1988) or the control plasmid p*GAL1-lacZ* (BOEKE *et al.* 1985) were introduced into isogenic strains JC344 (*HSX1*) and KK157 (*hsx1::LEU2*) that also contain the chromosomal Ty*1*m*his3*AI-270 element. Ty*1* transposition was induced by growing the cells on SC-ura (galactose) plates and spliced Ty*1*m*HIS3* transposition events were detected by replica plating onto SC-his-ura (glucose) plates (Figure 1). The *HSX1* strain DG1301 (containing the p*GAL1-lacZ* control plasmid) gave rise to a few transposition events, while the *HSX1* strain DG1302 (containing the pGTy*1*-H3*neo* helper plasmid) gave rise to hundreds of transposition events. In contrast, no Ty*1*m*HIS3* transposition events were present in the *hsx1::LEU2* strains DG1305 and DG1306, even though strain DG1306 contains a pGTy*1*-H3*neo* helper plasmid that was induced for transposition. Since the *hsx1::LEU2* mutation is recessive (Table 3), we showed that the pGTy*1*-H3*neo* helper plasmid is transposition-competent by testing pGTy*1*-H3*neo* transposition in an *hsx1::LEU2/HSX1* diploid strain (Table 4).

Several controls were performed to determine whether the *hsx1* mutation directly affected the Ty transposition process or whether the *hsx1* mutation affected RNA splicing or Ty RNA levels. The splicing efficiency of the Ty*1*m*his3*AI-270 transcript varied

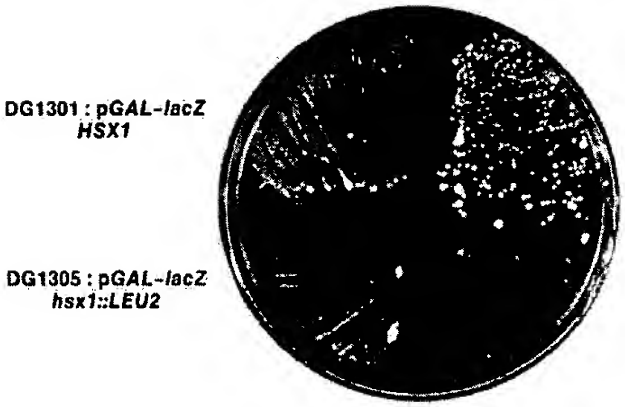


FIGURE 1.—Ty*1*m*his3*AI-270 transposition in an *hsx1* mutant background. Strains DG1301, DG1302, DG1305 and DG1306 contain the genomic Ty*1*m*his3*AI-270 element. The relevant plasmids and status of the *HSX1* gene are shown alongside the strains. These strains were tested for transposition by growing cells on SC-ura (galactose) plates for 7 days at 20°, replica plating to SC-his-ura (glucose), and incubating the replicas for 3 days at 30°.

TABLE 4

Ty*1neo* transposition in *hsx1/HSX1* diploid strains

Relevant genotype[a]	Relative transposition efficiency (%)[b]
HSX1/HSX1	42 (15/36)
hsx1::LEU2/HSX1	47 (16/34)

[a] Homozygous *HSX1/HSX1* diploids were obtained by mating strains DG1302 and KD198-16A. Heterozygous *hsx1::LEU2/HSX1* diploids were obtained by mating strains DG1306 and KD198-16A.
[b] In this transposition test, the transposition efficiency is the number of G418^r, Ura$^-$ plasmid segregants divided by the total number of Ura$^-$ plasmid segregants.

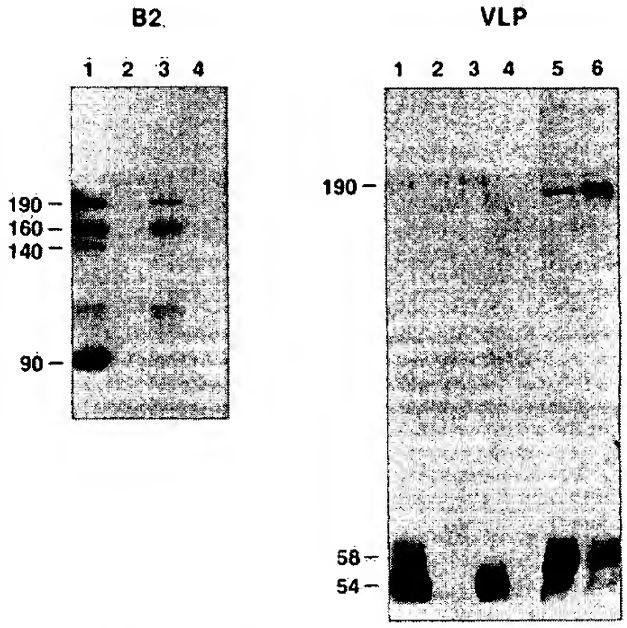


FIGURE 2.—Immunoblot analysis of Ty*1* proteins from an *hsx1* mutant background. Strains DG1302 (*HSX1*, pGTy*1*-H3*neo*; lane 1), DG1301 (*HSX1*, p*GAL-lacZ*; lane 2), DG1306 (*hsx1::LEU2*, pGTy*1*-H3*neo*; lane 3), DG1305 (*hsx1::LEU2*, p*GAL-lacZ*; lane 4), DG1333 (*HSX1*, pGTy*1*-H3*neo::SacI-1702*; lane 5), and DG1334 (*hsx1::LEU2*, pGTy*1*-H3*neo::SacI-1702*, lane 6) were induced for transposition by growth in SC-ura (galactose) medium and total protein was isolated for immunoblot analysis. Proteins were separated by electrophoresis on an SDS-8% polyacrylamide gel, transferred to a nitrocellulose membrane, and cross-reacted with B2 and VLP antisera. The B2 antiserum detects p90-Ty*1*-IN and its precursors. The VLP antiserum detects p54 and p58, which are VLP structural proteins derived from TYA1, as well as p190-TYA1-TYB1. The minor bands observed between p90-Ty*1*-IN and p140-TYB1 are probably caused by cellular proteolysis because they are present in immunoblots prepared from a Ty PR mutant (S. D. YOUNGREN and D. J. GARFINKEL, unpublished results). Ty*1* protein size estimates (in kilodaltons) are indicated.

from 12 to 20% in both *HSX1* or *hsx1::LEU2* strains as determined by RT-PCR. These splicing efficiencies agree with previous results where it was shown that about 12% of the Ty*1*m*his3*AI transposition events had lost the AI by splicing (CURCIO and GARFINKEL 1991). However, the overall Ty*1* and Ty*1*m*his3*AI-270 RNA levels were between 2- and 8-fold lower in an *hsx1::LEU2* mutant background when compared with *ACT1*, *PYK1* RNA or rRNA levels, although these differences were not completely reproducible.

To determine whether this moderate decrease in the level of Ty RNA could account for the more than 50-fold reduction in Ty*1* transposition, we assayed the level of pGTy*1*-H3*made2*AI retrotransposition (M. J. CURCIO and D. J. GARFINKEL, unpublished results) in an *hsx1::LEU2* mutant. In collateral experiments, the level of pGTy*1* expression in an *hsx1::LEU2* mutant was determined by immunoblotting (see below). The efficiency of Ty*1made2*AI transposition was reduced almost 70-fold in an *hsx1::LEU2* mutant background, while the level of *GAL1*-promoted Ty*1* proteins remained unchanged in the mutant (Figure 2). A similar decrease in transposition was also observed when an *HSX1* strain containing a pGTy*1*A-B*neo* plasmid with a mutation that corrects the frameshift was galactose-induced. Taken together, these results suggest that neither inhibition of splicing nor the lower concentration of chromosomal Ty*1* or Ty*1*m*his3*AI RNA can completely account for the reduction of Ty*1* transposition in an *hsx1::LEU2* mutant. Previous analyses have shown that increased expression of tRNA-Arg(CCU) (*HSX1*) negatively regulates Ty*1* transposition (XU and BOEKE 1990). Our results indicate that the *HSX1* gene is required for transposition of Ty*1* elements.

Mature TYB1 proteins do not accumulate in an *hsx1* disruption mutant: To further investigate the inhibition of Ty*1* transposition by *hsx1::LEU2*, we compared the levels and processing of Ty*1*-encoded proteins in isogenic *HSX1* and *hsx1* disruption strains (Figure 2). Total protein was isolated from strains DG1302 (*HSX1*, pGTy*1*-H3*neo*; lane 1), DG1301 (*hsx1*, p*GAL1-lacZ*; lane 2), DG1306 (*hsx1::LEU2*, pGTy*1*-H3*neo*; lane 3), and DG1305 (*hsx1::LEU2*, p*GAL1-lacZ*; lane 4) that were induced with galactose. The proteins were separated on SDS-polyacrylamide gels and immunoblotted. The resulting filters were reacted with B2 antiserum, which reacts with the full-length 190-kilodalton (kD) TYA1-TYB1 precursor protein, the 160-kD and 140-kD processing intermediates, and mature 90-kD Ty*1*-IN (GARFINKEL *et al.* 1991) or Ty*1*-VLP antiserum, which reacts with the 58-kD TYA1 precursor protein and the mature 54-kD TYA1 product (ADAMS *et al.* 1987; MULLER *et al.* 1987; YOUNGREN *et al.* 1988). Wild-type protein patterns were observed when the *HSX1* strain DG1302 was analyzed with B2 or Ty*1*-VLP antiserum (lane 1), or with an antiserum (B8) that detects p60-Ty*1*-RT/RH (B. FAIOLA and D. J. GARFINKEL, data not shown). As expected, strains DG1301 (lane 2) and DG1305 (lane 4) containing the heterologous expression plasmid p*GAL-lacZ* had very low levels of Ty*1* proteins (GARFINKEL *et al.* 1985; CURCIO and GARFINKEL 1992).

The *hsx1::LEU2* strain DG1305 (Figure 2, lane 3) displayed a different protein pattern when reacted with B2 and Ty*1*-VLP antisera. Essentially wild-type levels of the 190-kD TYA1-TYB1 precursor protein and 160-kD processing intermediate were detected using B2 antiserum. However, very little of the 140-kD precursor or 90-kD IN protein was detected. Similar results were obtained when an antiserum (B8) that detects RT/RH was used: the 190-kD and 160-kD TYB1 precursor proteins were present at wild-type levels, but the 140-kD precursor and the 60-kD Ty*1* RT/RH protein were barely detectable (B. FAIOLA and D. J. GARFINKEL, data not shown). When TYA1 proteins were analyzed with Ty*1*-VLP antiserum, normal levels of mature p54-TYA1 protein were observed in an *hsx1* mutant, but very little full-length p58-TYA1 precursor was detected even after extended exposure of the filter. Furthermore, similar protein patterns were observed when partially purified Ty*1*-VLPs were reacted with B2, B8, or Ty*1*-VLP antisera (B. FAIOLA and D. J. GARFINKEL, data not shown). These results suggest that the transposition defect observed in *hsx1* mutants is related to aberrant protein processing.

Ty*1* frameshifting increases in an *hsx1* disruption mutant: We tested whether the observed transposition defect in the *hsx1* mutant resulted from abnormal frameshifting using two different frameshifting assays. In the first assay, the *HSX1* strain KK242 and *hsx1::HIS3* mutant strain KK240 were transformed with pMB38-9merFusion and pMB38-9merWT plasmids in which the 0 (*TYA1*) and +1 (*TYA1-TYB1*) reading frames and *lacZ* are fused, respectively (Table 5). β-Galactosidase activity was determined from at least six different transformants of each plasmid and Ty*1* frameshifting efficiencies were calculated as described (see MATERIALS AND METHODS; BELCOURT and FARABAUGH 1990). A frameshifting efficiency of 35% was obtained in an *HSX1* background, which is comparable to published values (BELCOURT and FARABAUGH 1990). In contrast, the *hsx1::HIS3* disruption resulted in 91% frameshifting. The frameshifting efficiency was restored to 35% by a low copy number plasmid carrying the wild-type *HSX1* gene (pKK73; Table 6).

We also determined the Ty*1* frameshifting efficiency by quantitating the ratio of the unprocessed p58-TYA1 precursor to the p190-TYA1-TYB1 precursor in *HSX1* and *hsx1::LEU2* strains DG1333 and DG1334, respectively (Figure 2, lanes 5 and 6). To insure that unprocessed precursor proteins accumulated during the galactose induction, strains DG1333 and DG1334 contained a pGTy*1*-H3 plasmid with a well characterized Ty*1*-PR mutation, pGTy*1*-

TABLE 5

Translational frameshifting in an *hsx1* mutant

Relevant genotype	Frameshift site	β-Galactosidase units	Frameshifting efficiency (%)
HSX1	9merWT	2400	
	9merFusion	6800	35
	9merFusion(w/o AGG)	8900	27
hsx1::HIS3	9m34WT	5100	
	9merFusion	5600	91
	9merFusion(w/o AGG)	6100	84

Strains KK242 (*HSX1*) and KK240 (*hsx1::HIS3*) were transformed with plasmids pMB-9merWT, pMB38-9merFusion, and pMB38-9merFusion(w/o AGG). β-Galactosidase activities are the averages from six independent transformants. The frameshift efficiency is defined as the β-galactosidase activity of the 9merWT divided by the β-galactosidase activity of either the 9merFusion or the 9merFusion(w/o AGG) (BELLCOURT and FARABAUGH 1990).

TABLE 6

Translational frameshifting in an *hsx1* mutant KK240 containing plasmid copies of tRNA genes

Plasmid genotype	Frameshifting efficiency (%)
pKK73 (*HSX1*)	35
pKK74 [*hsx1*(*Mlu1**)]	98
pKK75 [*SUP201-0*(*UCU*)]	90
pKK76 [*SUP201-0-1*(*CCU*)]	65

Plasmids were introduced into strain KK240 (*hsx1::HIS3*) by transformation. Refer to Table 5 for experimental details.

H3*neo*::*SacI-1702* (YOUNGREN *et al.* 1988; GARFINKEL *et al.* 1991; CURCIO and GARFINKEL 1992). Proteins were analyzed by immunoblotting using Ty*1*-VLP antiserum, which recognizes TYA1 proteins and the 190-kD TYA1-TYB1 precursor protein (ADAMS *et al.* 1987; MULLER *et al.* 1987; YOUNGREN *et al.* 1988), and frameshifting efficiencies were calculated by densitometry (see MATERIALS AND METHODS).

The *HSX1* strain DG1333 (Figure 2, lane 5) showed the pattern of unprocessed 58-kDa and 190-kDa proteins expected from a Ty*1*-PR mutant (ADAMS *et al.* 1987; MULLER *et al.* 1987; YOUNGREN *et al.* 1988). A frameshifting efficiency of about 3% was obtained from densitometric scans of various exposures of the immunoblot. In contrast, the *hsx1::LEU2* strain DG1334 (Figure 2, lane 6) had much more of the 190-kD TYA1-TYB1 precursor and slightly less of the 58-kD TYA1 precursor than the *HSX1* parent strain DG1333 (Figure 2, lane 5). The *hsx1::LEU2* disruption mutant had a frameshifting efficiency of about 50%, which is about 17-fold higher than in an *HSX1* background. The overall level of Ty*1* protein also appeared to be similar in the *HSX1* or *hsx1* mutant backgrounds. These results suggest that the absence of tRNA-Arg(CCU) enhances ribosomal pausing at AGG and slippage of the leucyl-tRNA from CUU to UUA. Furthermore, the regulation of frameshifting by the *HSX1* gene is essential for Ty*1* transposition. The reduction in transposition in an *hsx1* mutant may be caused by a defect in protein processing that results from an aberrant stoichiometry of Ty proteins.

The capacity to translate an AGG codon does not limit β-galactosidase synthesis in an *hsx1* mutant: The *lacZ* fusion gene in the pMB38-9merFusion plasmid has only one AGG codon and it is located at the fusion site (BELCOURT and FARABAUGH 1990). That AGG codon is missing in the pMB38-9merFusion(w/oAGG) *lacZ* fusion gene. Therefore, the effect of a single AGG codon on β-galactosidase synthesis was determined in an *hsx1::HIS3* mutant. Interestingly, β-galactosidase activities in the *hsx1::HIS3* mutant or the *HSX1* parental strain harboring the pMB38-9merFusion and the pMB38-9merFusion(w/oAGG) plasmids were similar (Table 5). These results suggest that the capacity to translate the AGG codon does not limit β-galactosidase synthesis in an *hsx1* mutant. However, we do not know how the AGG is translated in an *hsx1* mutant. Since haploid cells contain more than eight tRNA-Arg(UCU) genes (BECKMANN, JOHNSON and ABELSON 1977), it is possible that tRNA-Arg(UCU) decodes AGG codons by near-cognate recognition when tRNA-Arg(CCU) is absent (YOKOYAMA *et al.* 1985).

Complementation of *hsx1* by a tRNA suppressor *SUP201-0-1(CCU)*: Although tRNA-Arg(UCU) may decode AGG codons, excess tRNA-Arg(UCU) does not inhibit frameshifting (BELCOURT and FARABAUGH 1990). This may be because of sequence or structural differences between tRNA-Arg(UCU) and tRNA-Arg(CCU) (Figure 3). Alternatively, the information needed to regulate Ty*1* frameshifting may reside within the anticodon. To determine if the CCU anticodon is sufficient to regulate Ty*1* transposition (Table 3) and frameshifting (Table 6), we constructed a low-copy-number plasmid carrying a mutant tRNA-Arg gene that has a CCU instead of a UCU anticodon. The *SUP201-0-1(CCU)* anticodon mutation was introduced into the *SUP201-0* tRNA-Arg(UCU) gene (THIREOS, PENN and GREER 1984; MORISHITA and UNO 1991), by oligonucleotide mutagenesis (refer to MATERIALS AND METHODS). Functionally active tRNAs were synthesized from these plasmids because a plasmid carrying the same 112-bp segment of DNA with a *SUP201* nonsense suppressor complemented the *cyr1-2* UGA allele (MORISHITA and UNO 1991; K. KAWAKAMI and Y. NAKAMURA, unpublished results).

To determine if *SUP201-0-1(CCU)* could suppress the transposition defect imposed by an *hsx1* mutation, strain KK157 containing Ty*1*m*his3*AI-270 and *hsx1::LEU2* was transformed with the suppressor plasmid pKK71 [*SUP201-0-1(CCU)*] or the parental plasmid pKK69 [*SUP201-0(UCU)*]. The level of Ty*1* transposition was partially restored when the pKK71 [*SUP201-0-1(CCU)*] plasmid was present in the *hsx1::LEU2* mutant (Table 3). This result suggests that the CCU anticodon can regulate transposition.

An *hsx1::HIS3* mutant strain KK240 harboring plasmids pMB38-9merWT or pMB38-9merFusion was transformed with plasmids pKK75 [*SUP201-0(UCU)*] and pKK76 [*SUP201-0-1(CCU)*] and frameshifting efficiencies were analyzed in these transformants (Table 6). The *SUP201-0-1(CCU)* mutant tRNA resulted in an intermediate level of frameshifting. Interestingly, frameshifting in the pKK76 [*SUP201-0-1(CCU)*] transformant was higher (65%) than in the pKK73 [*HSX1*; tRNA-Arg(CCU)] transformant (35%). This result is consistent with the lower level of transposition of the pKK71 [*SUP201-0-1(CCU)*] transformant (1.1×10^{-6}) when compared to the pKK67 [*HSX1*; tRNA-Arg(CCU)] transformant (6.4×10^{-6}; Table 3). Therefore, although *SUP201-0-1(CCU)* can partially regulate Ty transposition and frameshifting, it does not work as well as tRNA-Arg(CCU) encoded by *HSX1*. Other aspects of *SUP201-0-1(CCU)* expression or structure may prevent full complementation of the *hsx1* mutation. These results also suggest that base pairing at the third position of the second codon in the frameshift heptamer is essential for regulating Ty*1* transposition and frameshifting.

Increasing TYA1 expression restores Ty*1* transposition in an *hsx1* mutant: Our results indicate that more of TYA1-TYB1 fusion protein is translated in

```
SUP201-0  GCUCGCGUGGGCGUAAUGGCAACGCGUCUGACUUCUAAUCAGAAGAUUAUGGGUUCGACCCCCAUCGUGAGUG
          * **  ************ ********** ** **** ****** * ********* *** *  **  *
HSX1      GUUCCGUUGGCGUAAUGGUAACGCGUCUCCCCUCCUAAGGAGAAGACUGCGGGUUCGAGUCCCGUACGGAACG
                              MluI
```

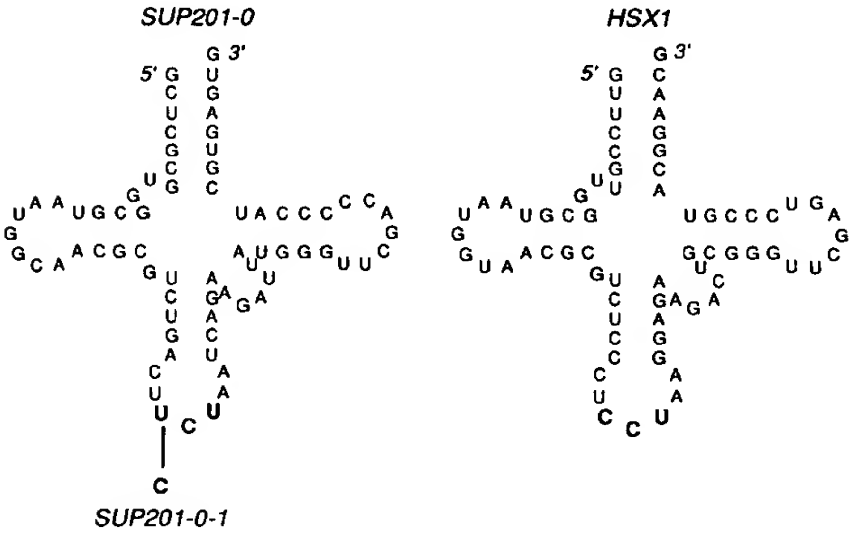


FIGURE 3.—Comparison of *SUP201-0* and *HSX1*. Nucleotide sequences of the *SUP201-0* and *HSX1* tRNAs are shown. Identical nucleotides are indicated by asterisks. Anticodons are in bold lettering. The clover leaf structure of the *SUP201-0* and *HSX1* genes, and the *SUP201-0-1(CCU)* mutation are also shown.

the *hsx1* disruption mutant, and that this altered stoichiometry of TYA1 to TYA1-TYB1 precursor proteins may inhibit Ty*1* transposition. Therefore, providing more TYA1 protein should restore Ty*1* transposition in an *hsx1* mutant background. To test this idea, the efficiency of Ty*1* transposition was determined with strains DG1344 (*HSX1*) and DG1347 (*hsx1::LEU2*) containing the helper plasmid pGTyA*1neo* (*PGK1* ter.), and strains DG1301 (*HSX1*) and DG1305 (*hsx1::LEU2*) containing the heterologous expression plasmid p*GAL-lacZ*. The helper plasmid pGTyA*1neo* (*PGK1* ter.) contains a complete *TYA1* gene, about 25 codons of N-terminal *TYB1* sequence, the *neo* marker gene, and a transcriptional terminator from the *PGK1* gene in place of the downstream long terminal repeat. Liquid cultures of strains DG1301, DG1305, DG1344 and DG1347 were galactose-induced and transpositions of the chromosomal Ty*1*m*his3*AI-270 element were selected on SC-his-ura (glucose) medium (Table 7). Although expression of the helper plasmid pGTyA1*neo* (*PGK1* ter.) did not markedly affect Ty*1* transposition in an *HSX1* strain, expression of *TYA1* stimulated Ty*1* transposition 50-fold more in an *hsx1* strain than in an *HSX1* strain. In addition, galactose-induction of pGTyA*1neo* (*PGK* ter.) did not affect the level of full-length Ty*1* RNA in an *HSX1* or *hsx1* disruption strain (B. FAIOLA and D. J. GARFINKEL, data not shown). As expected, the level of the TyA*1neo* (*PGK* ter.) transcript was the same in the *HSX1* and *hsx1::LEU2* mutant strains. These results indicate that expression of pGTyA*1neo* (*PGK1* ter.) stimulates Ty*1* transposition in an *hsx1* mutant background by restoring the proper stoichiometry of TYA1 to TYA1-TYB1 precursor proteins.

TABLE 7

Effect of pGTyA*1neo*(*PGK1* ter.) expression on Ty*1*m*his3*A1-270 transposition in an *hsx1* mutant

Strain	Relevant genotype	Relative transposition efficiency
DG1301	*HSX1*, p*GAL1-lacZ*	1.3×10^{-7}
DG1344	*HSX1*, pGTyA*1neo*(*PGK1* ter.)	4.8×10^{-7}
DG1305	*hsx1::LEU2*, p*GAL1-lacZ*	$<3.7 \times 10^{-9}$
DG1347	*hsx1::LEU2*, pGTyA*1neo*(*PGK1* ter.)	2.4×10^{-5}

These strains contain the genomic Ty*1*m*his3*AI-270 element. Relative transposition efficiencies were determined from liquid cultures grown in SC-ura (galactose) as described in MATERIALS AND METHODS. The relative transposition efficiency is the number of His^+, Ura^+ colonies divided by the number of Ura^+ colonies. Each measurement represents the mean of four cultures. The total number of colony-forming units was similar within each set of cultures. Refer to Table 2 for more information.

DISCUSSION

Our study reveals that the *HSX1* gene is necessary for Ty*1* transposition because elimination of this gene causes a significant transposition defect. Our work also shows that Ty*1* translational frameshifting increases dramatically in an *hsx1* disruption mutant. The *hsx1* mutant defects in frameshifting and Ty*1* transposition are completely complemented by the wild-type *HSX1* gene and partially complemented by the

mutant *SUP201-0-1(CCU)* gene, while no complementation occurs with the *SUP201-0(UCU)* gene (Tables 3 and 6). Therefore, at least some of the information required for Ty*1* frameshifting is provided by the CCU anticodon. The partial complementation activity of the mutant *SUP201-0-1* tRNA-Arg(CCU) suggests two possibilities. First, *SUP201-0-1(CCU)* may be expressed at a lower level than *HSX1*, thus directly affecting the level of tRNA-Arg(CCU) available for frameshifting. Second, *SUP201-0-1(CCU)* may not recognize the AGG codon within the context of the frameshift heptamer as well as *HSX1*, since the *SUP201-0-1 and HSX1* tRNA genes differ by 20 nucleotide changes (Figure 3). RAFTERY and YARUS (1987) have shown that the structure of the proximal anticodon stem affects efficiency of a tRNA suppressor of *E. coli* and suggested that it is a part of the extended anticodon. The 2-bp difference in the anticodon stem between *SUP201-0-1* and the *HSX1* tRNAs may result in the altered AGG codon recognition activity of the *SUP201-0-1* tRNA.

Both −1 and +1 frameshifting mechanisms used by a variety of RNA viruses, retroviruses, and retrotransposons apparently require a translational pause for optimum efficiency (reviewed by Hatfield *et al.* 1992). For example, the translational pause in retroviral −1 frameshifting is created by a pseudoknot located a few nucleotides downstream of the frameshift, whereas Ty*1* +1 frameshifting uses a the rare tRNA-Arg(CCU). Our results are consistent with the +1 frameshifting model proposed by BELCOURT and FARABAUGH (1990). According to this model, the increase in +1 frameshifting results from a longer translational pause in an *hsx1* mutant created by the absence of tRNA-Arg(CCU). The longer translational pause regulates translation of *TYB1-pol* by allowing more time for the tRNA-Leu to slip from the 0-frame CUU codon in *TYA1* to the +1-frame UUA codon in *TYB1*.

Two different approaches were used to estimate the increase in frameshifting that occurs in an *hsx1* disruption mutant. First, frameshifting was measured using the minimal heptamer sequence with *lacZ* as a reporter gene (BELCOURT and FARABAUGH 1990). The absence of tRNA-Arg(CCU) increased frameshifting as measured by β-galactosidase activity about 3-fold. Second, frameshifting was measured by immunoblotting using Ty*1*-VLP antiserum and a Ty*1*-PR mutant defective in protein processing. The increase in frameshifting at the **CUU-AGG-C** sequence leads to accumulation of slightly less p54-TYA1 protein and much more p190-TYA1-TYB1 fusion protein. Using this assay, frameshifting increased about 17-fold in an *hsx1* background.

We estimate that Ty*1* frameshifting occurs at about a 3% efficiency in an *HSX1* background by immunoblotting. In other words, 3% of ribosomes translating the *TYA1-gag* open reading frame undergo a +1 frameshift and continue translating the *TYB1-pol* open reading frame. It is somewhat surprising that the Ty*1* frameshifting efficiency of 3% is about 5–10-fold lower than that obtained by lacZ fusion analysis. It is possible that we have underestimated the Ty*1* frameshifting efficiency obtained from immunoblotting because of an inability to detect the p190-TYA1-TYB1 precursor protein. However, control experiments suggest that p58-TYA1 and p190-TYA1-TYB1 are transferred at about the same rate under the immunoblotting conditions used in this study, bind to Ty*1*-VLP antiserum with comparable affinities, and have similar turnover rates (A.-M. HEDGE and D. J. GARFINKEL, unpublished results; CURCIO and GARFINKEL 1992). There may also be differences in translation rates of *lacZ* in yeast, or in the placement of the frameshift heptamer relative to the start of translation that contribute to this apparent discrepancy (P. J. FARABAUGH, unpublished results).

The Ty*1* frameshifting efficiency of 3% obtained by immunoblot analysis is comparable to the efficiencies obtained from several viral systems that utilize different mechanisms for translation of the pol gene. Retroviruses that utilize programmed ribosomal frameshifting or read-through suppression undergo translational suppression at an efficiency of about 5% (reviewed by HATFIELD *et al.* 1992). Yeast Ty*3* retrotransposons have a +1 frameshifting efficiency of about 4%, even though these elements use a different frameshifting site (KIRCHNER, SANDMEYER and FORREST 1992) and mechanism than Ty*1* or Ty2 elements (P. J. FARABAUGH, unpublished results). In addition, the yeast L-A double-stranded RNA virus undergoes −1 frameshifting to express its *pol* gene at an efficiency of about 2% (DINMAN, ICHO and WICKNER 1991). Even though the molecular mechanisms underlying these expression strategies are quite different, a certain ratio of "structural" (Gag) proteins to "catalytic" (Gag-Pol) proteins may be a general requirement for formation of a transposition/replication-competent particle.

Immunoblot analysis suggested that a processing defect of the TYA1-TYB1 fusion protein is related to the lower level of Ty*1* transposition in an *hsx1* disruption mutant. The protein cleavages required to form p54-TYA1 and the p160 processing intermediate still occur, while the proteolytic cleavage required to convert the p160 processing intermediate to p23-PR and the p140 processing intermediate apparently do not (Figure 4). Since formation of p140-TYB1 is defective, it follows that low amounts of mature IN and RT/RH are detected in an *hsx1::LEU2* mutant. Perhaps Ty*1*-PR is not completely activated when more of the TYA1-TYB1 fusion protein is produced. Alternatively, normal amounts of p140, IN and RT/RH

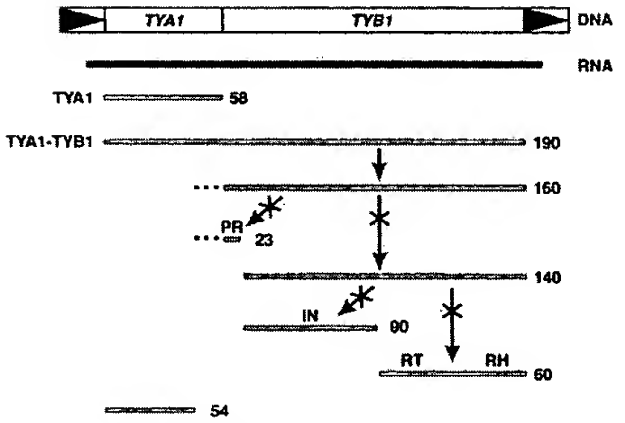


FIGURE 4.—Scheme for TYB1 protein processing in an *hsx1* mutant (modified from GARFINKEL *et al.* 1991). p190-TYA1-TYB1, the 190-kD product of the *gag-TYA1* and *pol-TYB1* genes, is cleaved near the frameshift region (the vertical line separating *TYA1* and *TYB1*). This proteolytic cleavage releases p160-TYB1, which is normally cleaved to form Ty*1*-PR (23 kD) and p140-TYB1. Cleavage of p140-TYB1 produces IN (90 kD) and RT/RH (60 kD). The dotted lines indicate that p160 and p23 may be encoded by both *TYA1* and *TYB1*. The arrows show that neither the p140-TYB1 percursor nor mature p90-Ty*1*-IN and p60-Ty*1*-RT/RH accumulate in an *hsx1* mutant. Also shown is the p58-TYA1 precursor and p54 processed product, which are the major structural components of Ty*1*-VLPs. In an *hsx1* mutant, we detect p54-TYA1 but not the p58-TYA1 precursor.

may be synthesized, but are rapidly degraded because of an *hsx1*-dependent defect in Ty*1*-VLP assembly.

To prove that aberrant protein stoichiometry is the major reason for the block in Ty*1* transposition in an *hsx1* disruption mutant, we showed that a pGTy*1* plasmid expressing just the *TYA1* gene not only restores Ty*1* transposition in an *hsx1::LEU2* mutant, but stimulates transposition to a level 50-fold higher than is observed in an *HSX1* strain. We also showed that overexpression of *TYA1* does not alter the level of Ty*1* RNA in an *hsx1* mutant. These results suggest that overexpression of *TYA1* enhances the utilization of Ty*1* RNA as a transposition template by rebalancing the level of TYA1 and TYA1-TYB1 proteins required to make transposition-competent Ty*1*-VLPs in an *hsx1* mutant, even though the absolute level of Ty*1* RNA is somewhat lower in the *hsx1* mutant. Furthermore, since *GAL1*-promoted Ty*1* transposition decreases about 70-fold without a concomitant decrease in *GAL1*-promoted Ty*1* protein levels in an *hsx1* disruption mutant, whatever effect the *hsx1* mutation has on Ty*1* RNA levels is limited to chromosomal Ty*1* elements. These results suggest that the *hsx1* mutation may affect chromosomal Ty*1* RNA accumulation, but we have not investigated this idea further.

The stimulation of Ty transposition that occurs in an *hsx1* mutant when *TYA1* is overexpressed supports and extends previous biochemical and genetic studies that identified the availability of Ty*1*-PR, which is encoded by *TYB1*, as a rate-limiting step in the Ty*1* retrotransposition cycle (CURCIO and GARFINKEL 1992). Since more TYA1-TYB1 precursor protein is made in an *hsx1* mutant, the availability of TYA1 protein becomes rate-limiting under these conditions. Therefore, a specific ratio of TYA1 to TYA1-TYB1 precursor proteins is required to form fully processed Ty*1* proteins and functional Ty*1*-VLPs.

Several retrovirus, retrotransposon and endogenous viral mutants in which *gag* and *pol* have been artificially fused are defective in particle formation, replication and infectivity. For example, fusion of *gag* and *pol* genes blocks production of infectious Moloney murine leukemia virus (FELSENSTEIN and GOFF 1988) and human immunodeficiency virus (PARK and MORROW 1992). In Moloney murine leukemia virus, the Gag-Pol precursor protein is produced, but neither protein processing nor particle formation occurs. In human immunodeficiency virus, the Gag-Pol protein is produced and processed, but particles do not form. A protein processing and transposition defect similar to the one created in an *hsx1* mutant is observed when *TYA1* and *TYB1* are fused by deleting one base at the frameshift site of a pGTy*1* plasmid and transposition is galactose-induced in an *HSX1* strain. Preliminary experiments suggest that Ty*1*-VLPs are formed in an *hsx1* mutant (B. FAIOLA and D. J. GARFINKEL, unpublished results) or when just the TYA1-TYB1 fusion protein is expressed (J. D. BOEKE and D. J. GARFINKEL, unpublished results). Recently, a Ty*3* *GAG3-POL3* fusion mutant has been analyzed for defects in transposition and Ty*3*-VLP formation using a pGTy*3* expression system (KIRCHNER, SANDMEYER and FORREST 1992). The fusion mutant is transposition-defective, but can be rescued by coexpression of *GAG3* or just the capsid domain of *GAG3*. Protein processing of GAG3 capsid protein and Ty*3*-IN is altered in the mutants, as is individual Ty*3* protein and Ty*3*-VLP yield. Optimal ribosomal frameshifting and the proper Gag to Gag-Pol protein ratio are also required for L-A virus propagation in yeast (DINMAN and WICKNER 1992). Therefore, Ty*1* and Ty*3* elements seem to be unique in that some particle assembly can take place when excess Gag-Pol precursor protein is synthesized (KIRCHNER, SANDMEYER and FORREST 1992) when only Gag protein is synthesized (ADAMS *et al.* 1987; BURNS *et al.* 1992), or when PR-dependent protein processing is blocked (ADAMS *et al.* 1987; MULLER *et al.* 1987; YOUNGREN *et al.* 1988; KIRCHNER and SANDMEYER 1993).

In summary, our work has identified an essential role for *HSX1* in Ty*1* frameshifting and transposition. This is one of a small but growing collection of cellular genes required for Ty*1* transposition that act post-transcriptionally (reviewed by BOEKE and CHAPMAN 1991; GARFINKEL 1992). The additional defects of an

hsx1 disruption mutant (KAWAKAMI *et al.* 1992) may allow us to select second-site suppressors that restore Ty*1* transposition without affecting Ty*1* frameshifting mediated by tRNA-Arg. These suppressors may identify additional cellular genes involved in Ty*1* frameshifting or Ty*1*-VLP assembly.

We thank M. J. CURCIO, K. J. DURBIN, Y. MATSUI, T. MORISHITA and P. ROGAN for plasmids and strains. We thank S. BROWN, D. COURT, M. J. CURCIO and A. REIN for their helpful comments and J. HOPKINS for preparing the manuscript. This work was supported in part by The Ministry of Education, Science and Culture, Japan (K.K. and Y.N.); U.S. Public Health Service grants GM 29480 (S.P. and P.J.F.) and GM 36481 (D.P.M. and J.D.B.); and the National Cancer Institute under contract N01-CO-74101 with ABL (B.F., D.J.G., and J.N.S.). The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

LITERATURE CITED

ADAMS, S. E., J. MELLOR, K. GULL, R. B. KIM, M. F. TUITE, S. M. KINGSMAN and A. J. KINGSMAN, 1987 The functions and relationships of Ty-VLP proteins in yeast reflect those of mammalian retroviral proteins. Cell **49:** 111–119.

BECKMANN, J. S., P. F. JOHNSON and J. ABELSON, 1977 Cloning of yeast transfer RNA genes in *Escherichia coli*. Science **196:** 205–208.

BELCOURT, M. F., and P. J. FARABAUGH, 1990 Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. Cell **62:** 339–352.

BOEKE, J. D., and K. B. CHAPMAN, 1991 Retrotransposition mechanisms. Curr. Opin. Cell Biol. **3:** 502–507.

BOEKE, J. D., D. J. GARFINKEL, C. A. STYLES and G. R. FINK, 1985 Ty elements transpose through an RNA intermediate. Cell **40:** 491–500.

BOEKE, J. D., and S. B. SANDMEYER, 1991 Yeast transposable elements, pp. 193–262 in *The Molecular and Cellular Biology of the Yeast Saccharomyces: Genome Dynamics, Protein Synthesis, and Energetics*, Vol. 1, edited by J. R. BROACH, J. R. PRINGLE and E. W. JONES. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

BOEKE, J. D., H. XU and G. R. FINK, 1988 A general method for the chromosomal amplification of genes in yeast. Science **239:** 280–282.

BURNS, N. R., H. R. SAIBIL, N. S. WHITE, J. F. PARDON, P. A. TIMMINS, S. MARK, H. RICHARDSON, B. M. RICHARDS, S. E. ADAMS, S. M. KINGSMAN and A. J. KINGSMAN, 1992 Symmetry, flexibility and permeability in the structure of yeast retrotransposon virus-like particles. EMBO J. **11:** 1155–1164.

CLARE, J. J., M. BELCOURT and P. J. FARABAUGH, 1988 Efficient translational frameshifting occurs within a conserved sequence of the overlap between the two genes of a yeast Ty*1* transposon. Proc. Natl. Acad. Sci. USA **85:** 6816–6820.

CLARE, J. J., and P. J. FARABAUGH, 1985 Nucleotide sequence of a yeast Ty element: Evidence for an unusual mechanism of gene expression. Proc. Natl. Acad. Sci. USA **82:** 2829–2833.

CURCIO, M. J., and D. J. GARFINKEL, 1991 Single-step selection for Ty*1* element retrotransposition. Proc. Natl. Acad. Sci. USA **88:** 936–940.

CURCIO, M. J., and D. J. GARFINKEL, 1992 Posttranslational control of Ty*1* retrotransposition occurs at the level of protein processing. Mol. Cell. Biol. **12:** 2813–2825.

CURCIO, M. J., N. J. SANDERS and D. J. GARFINKEL, 1988 Transcriptional competence and transcription of endogenous Ty elements in *Saccharomyces cerevisiae*: Implications for regulation of transposition. Mol. Cell. Biol. **8:** 3571–3581.

DINMAN, J. D., T. ICHO and R. B. WICKNER, 1991 A −1 ribosomal frameshift in a double-stranded RNA virus of yeast forms a gag-pol fusion protein. Proc. Natl. Acad. Sci. USA **88:** 174–178.

DINMAN, J. D., and R. B. WICKNER, 1992 Ribosomal frameshifting efficiency and *gag/gag-pol* ratio are critical for yeast M_1 double-stranded RNA virus replication. J. Virol. **66:** 3669–3676.

EICHINGER, D. J., and J. D. BOEKE, 1988 The DNA intermediate in yeast Ty*1* element transposition copurifies with virus-like particles: cell-free Ty*1* transposition. Cell **54:** 955–966.

FELSENSTEIN, K., and S. GOFF, 1988 Expression of the gag-pol fusion protein of Moloney murine leukemia virus without gag protein does not induce virion formation or proteolytic processing. J. Virol. **62:** 2179–2182.

GAFNER, J., E. M. DE ROBERTIS and P. PHILIPPSEN, 1983 δ sequences in the 5′ non-coding region of yeast tRNA genes. EMBO J. **2:** 583–591.

GARFINKEL, D. J., 1992 Retroelements in microorganisms, pp. 107–158 in *The Retroviridae*, Vol. 1, edited by J. A. LEVY. Plenum Press, New York.

GARFINKEL, D. J., J. D. BOEKE and G. R. FINK, 1985 Ty element transposition: reverse transcription and virus-like particles. Cell **42:** 507–517.

GARFINKEL, D. J., A.-M. HEDGE, S. D. YOUNGREN and T. D. COPELAND, 1991 Proteolytic processing of *pol-TYB* proteins from the yeast retrotransposon Ty*1*. J. Virol. **65:** 4573–4581.

HATFIELD, D. L., J. G. LEVIN, A. REIN and S. OROSZLAN, 1992 Translational suppression in retroviral gene expression. Adv. Virus Res. **41:** 193–239.

ITO, H., Y. FUKUDA, K. MURATA and A. KIMURA, 1983 Transformation of intact yeast cells treated with alkali cations. J. Bacteriol. **153:** 163–168.

KAWAKAMI, K., B. K. SHAFER, D. J. GARFINKEL, J. N. STRATHERN and Y. NAKAMURA, 1992 Ty element-induced temperature-sensitive mutations of *Saccharomyces cerevisiae*. Genetics **131:** 821–832.

KIRCHNER, J., and S. B. SANDMEYER, 1993 Proteolytic processing of Ty*3* proteins is required for transposition. J. Virol. **67:** 19–28.

KIRCHNER, J., S. B. SANDMEYER and D. B. FORREST, 1992 Transposition of a Ty*3 GAG3-POL3* fusion mutant is limited by availability of capsid protein. J. Virol. **66:** 6081–6092.

MANDECKI, W., 1986 Oligonucleotide-directed double-strand break repair in plasmids of *Escherichia coli*: a method for site-specific mutagenesis. Proc. Natl. Acad. Sci. USA **83:** 7177–7181.

MORISHITA, T., and I. UNO, 1991 Genetic and molecular analyses of the *SUP201* gene: a tRNA-Arg nonsense suppressor of yeast *cyr1-2*. J. Bacteriol. **173:** 2406–2408.

MULLER, F., K.-H. BRUHL, K. FREIDEL, K. V. KOWALLIK and M. CIRIACY, 1987 Processing of Ty proteins and formation of Ty virus-like particles in *Saccharomyces cerevisiae*. Mol. Gen. Genet. **207:** 421–429.

PARK, J., and C. D. MORROW, 1992 Overexpression of the *gag-pol* precursor from human immunodeficiency virus type 1 proviral genomes results in efficient proteolytic processing in the absence of virion production. J. Virol. **65:** 5111–5117.

RAFTERY, L. A., and M. YARUS, 1987 Systematic alterations in the anticodon arm make tRNA-Glu-Su-ochre a more efficient suppressor. EMBO J. **6:** 1499–1506.

ROSE, M. D., F. WINSTON, and P. HIETER, 1990 *Laboratory Course Manual for Methods in Yeast Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

ROSE, M. D., P. NOVICK, J. H. THOMAS, D. BOTSTEIN and G. R.

FINK, 1987 A *Saccharomyces cerevisiae* genomic plasmid bank based on a centromere-containing shuttle vector. Gene **60:** 237–243.

ROTHSTEIN, R., 1991 Targeting, disruption, replacement, and allele rescue: Integrative DNA transformation in yeast. Methods Enzymol. **194:** 281–301.

SAIKI, R., S. SHARF, F. FALOONA, K. B. MULLIS, G. T. HORN, H. A. ERLICH and N. ARNHEIM, 1985 Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science **230:** 1350–1354.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning*, Ed. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SIKORSKI, R. S., and P. HIETER, 1989 A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics **122:** 19–27.

THIREOS, G., M. D. PENN and H. GREER, 1984 5′-Untranslated sequences are required for the translational control of a yeast regulatory gene. Proc. Natl. Acad. Sci. USA **81:** 5096–5100.

WANG, A. M., M. V. DOYLE and D. F. MARK, 1989 Quantitation of mRNA by the polymerase chain reaction. Proc. Natl. Acad. Sci. USA **86:** 9717–9721.

XU, H., and J. D. BOEKE, 1990 Host genes that influence transposition in yeast: The abundance of a rare tRNA regulates Ty*1* transposition frequency. Proc. Natl. Acad. Sci. USA **87:** 8360–8364.

YOKOYAMA, S., T. WATANABE, K. MURAO, H. ISHIKURA, Z. YAMAIZUMI, S. NISHIMURA and T. MIYAZAWA, 1985 Molecular mechanism of codon recognition by tRNA species with modified uridine in the first position of the anticodon. Proc. Natl. Acad. Sci. USA **82:** 4905–4909.

YOUNGREN, S. D., J. D. BOEKE, N. J. SANDERS and D. J. GARFINKEL, 1988 Functional organization of the retrotransposon Ty from *Saccharomyces cerevisiae*. Mol. Cell. Biol. **8:** 1421–1431.

Communicating editor: A. G. HINNEBUSCH

EXHIBIT G

J Bacteriol, February 1998, p. 840-845, Vol. 180, No. 4
0021-9193/98/$04.00+0

$tRNA^{Arg}$ (*fimU*) and Expression of SEF14 and SEF21 in *Salmonella enteritidis*

S. C. Clouthier,[1,2] S. K. Collinson,[1] A. P. White,[1] P. A. Banser,[1] and W. W. Kay[1,2,*]

Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia V8W 3P6,[1] and Microtek International Ltd., Saanichton, British Columbia V8X 3X1,[2] Canada

Received 21 August 1997/Accepted 16 December 1997

ABSTRACT

A Tn*10* insertion affecting SEF14 fimbrial synthesis in *Salmonella enteritidis* was located 13 bp upstream of a gene designated *fimU*. The 77-bp DNA sequence of *fimU* from *S. enteritidis* was identical to that of *fimU* encoding $tRNA^{Arg}$ (UCU) from *Salmonella typhimurium* and 96% identical to that of the *Escherichia coli argU* homolog. Furthermore, the open reading frame adjacent to and overlapping the 3' end of *fimU* was similar to the prophage DLP12 integrase gene. The *fimU*-encoded transcript comigrated with total cellular tRNA and was predicted to form a tRNA-like cloverleaf structure containing the arginine anticodon UCU. Thus, *fimU* encoded a $tRNA^{Arg}$ specific for the rare codon AGA. *fimU* mapped to the SEF21 *fim* operon located 15 C's from the *sefI4* gene cluster. Although *fimU* was located within the SEF21 *fim* gene cluster, the *fimU* Tn*10* insertion mutant of *S. enteritidis* was found to be defective in SEF14 as well as SEF21 (type 1) fimbria production. SEF17 and SEF18 fimbria production was not affected. Complementation of this mutant with plasmid-borne *fimU* restored normal production of the fimbrins SefA and FimA as well as their respective fimbriae SEF14 and SEF21. This is the first description of tRNA simultaneously controlling the production of two distinct fimbriae.

INTRODUCTION

Regulation of fimbria biosynthesis in bacteria is multifactorial and complex. In *Escherichia coli*, the expression of type 1 fimbriae is transcriptionally regulated in part by an inversion-dependent, phase-variable mechanism that

▼ Results
▼ Discussion
▼ References

involves two site-specific recombinases (17, 24, 27) and a $tRNA^{Leu}$ molecule (32). $tRNA^{Leu}$, specific for the rare leucine codon UUG, stimulates type 1 fimbria synthesis by influencing the switch from phase off to phase on (35). Recently, type 1 fimbria expression in *Salmonella typhimurium* has been shown to be regulated by mechanisms that are different from those controlling type 1 fimbria expression in *E. coli* (41). However, a common regulatory theme does exist in that a tRNA, specific for the rare arginine codons AGA and AGG, is required (40). Swenson et al. (40) suggest that the amount of $tRNA^{Arg}$ (UCU) available in *S. typhimurium* may influence the expression of three genes encoding regulatory proteins of the *fim* gene cluster, since in each of these genes there is a high frequency of rare AGA codons recognized by $tRNA^{Arg}$ (UCU).

Salmonella enteritidis 27655-3b produces at least four fimbrial types: SEF17 (10), SEF18 (6), SEF21 (type 1 fimbriae) (30), and SEF14 (7, 14). Although little is known about how the expression of the operons is regulated, SEF21 and SEF14 fimbriae are produced under similar environmental conditions (5, 12). Thus, the question arises as to whether or not their expression is coregulated. In a previous study, a Tn*10* insertion mutant, *S. enteritidis* 3b-122, was generated which no longer produced SEF14 fimbriae and carried the transposon outside of *sefA*, the structural gene for these fimbriae (14). Further characterization of 3b-122 in this study indicated that this mutant was also defective in type 1 fimbria (SEF21) production, suggesting that the Tn*10* interrupted a gene whose product coregulated the expression of both SEF14 and SEF21 fimbriae. The results of this study show for the first time that the production of two fimbriae is coregulated by the same tRNA.

MATERIALS AND METHODS

▲ Top
▲ Abstract
▲ Introduction
• Materials & Methods
▼ Results
▼ Discussion
▼ References

Bacterial strains, plasmids, media, and growth conditions. *S. enteritidis* 27655-3b, originally isolated from human feces, was provided by T. Wadstrom (University of Lund, Lund, Sweden). *S. enteritidis* 27655-3b-122, a Tn*10* insertion mutant of the parent strain, was constructed by Feutrier et al. (14). *E. coli* DH5α and *S. enteritidis* 3b-122 were used as hosts for pSFA (11), pLU/TA 4-1, and pGEM-T1. To create pLU/TA 4-1, PCR-amplified *fimU* was cloned into pGEM-T (Promega Corp.), a TA cloning vector containing 3'-terminal thymidines. To create pGEM-T1, the 3'-overhanging thymidines of pGEM-T were filled in with dATP and T4 DNA polymerase prior to ligation (38).

Bacteria were grown at 37°C with shaking in Luria-Bertani (LB) broth (36) supplemented with ampicillin to a final concentration of 250 μg/ml except where noted. To analyze the production of fimbriae by *S. enteritidis*, the cells were grown in various liquid media under different growth conditions (Table 1). Cultures grown in LB broth and terrific broth (TFB) (38) were transferred to ice 24 h after inoculation, whereas cultures grown in colonization factor antigen (CFA) medium (13) and T broth (10) were transferred to ice 48 h after inoculation. All the cultures were standardized to an optical density at 630 nm (OD_{630}) of 1.

TABLE 1. Production of SefA, FimA, AgfA, and SefD fimbrins by

View this table: *S. enteritidis* 3b and *S. enteritidis* 3b-122
[in this window]
[in a new window]

Subcloning Tn*10* from *S. enteritidis* 27655 3b-122. *S. enteritidis* 3b-122 chromosomal DNA was isolated by the method of Alm et al. (1), purified by CsCl centrifugation (38), and digested with *Hin*dIII. To subclone the Tn*10*-containing chromosomal DNA fragment, size fractionated *Hin*dIII fragments (2 to 3 and 3 to 5 kb) were purified from an agarose gel with Sephaglas (Pharmacia Biotech), ligated to *Hin*dIII-digested and -dephosphorylated cloning vector pTZ19R, and then introduced into *E. coli* DH5α by transformation (38). A total of 2,880 colonies grown on Hybond N^+ membranes (Amersham) were screened by hybridization to the oligonucleotide probe Tn*10* IS10L+R (5' GCAGAATTGGTAAAGAGA 3'). This probe, complementary to the sequence located 134 bp inside the insertion sequence of Tn*10*, was used to identify Tn*10*-containing clones. The probe, end labelled with [γ-^{32}P]ATP, was hybridized to the membranes at 45°C in prehybridization buffer (38) containing 200 μg of herring sperm DNA (Sigma)/ml. Following hybridization, the membranes were washed in 0.2× SSC buffer (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) containing 0.1% sodium dodecyl sulfate (SDS) at 45°C, and the results were recorded by autoradiography on Kodak BioMax film.

DNA sequencing and computer analyses. The Tn*10*-positive clones and the three *fimU* PCR products amplified with primers located outside the *fimU* gene were sequenced with Sequenase version 2 (United States Biochemicals). The custom oligonucleotide primer Tn*10*IS10L+R was synthesized on a PCR-MATE EP model 391 DNA synthesizer (Applied Biosystems Inc.). The DNA sequences obtained were analyzed with DNA Strider 1.1 (26). Similarity searches of the National Center for Biotechnology Information (NCBI) databases were conducted with the program BLASTN (2).

PCR amplification of *fimU*. Custom oligonucleotide primers *fimU*LT (TAATAGCGATACGCAGAATTCAAAAATATCCTACACGGCAGG) and *fimU*LB (CAGATATGCTCACCTAAGCTTTAATCATTTAACGGAACACGG) were designed based on the *S. typhimurium* chromosomal DNA sequence flanking *fimU* and were synthesized by Gibco BRL. *fimU* was PCR amplified from a previously prepared cosmid clone, pPB523 (12), with *fimU*LT and *fimU*LB. To facilitate the cloning of the amplified product, the primers were designed to contain an *Eco*RI site and a *Hin*dIII site, respectively (underlined). Amplification was carried out in a 100-μl reaction volume containing 10 μl of pPB523 (0.01 μg/ml), 25 pmol of each primer, the four deoxynucleotide triphosphates (Boehringer Mannheim) at 0.5 mM each, and 2 U of *Taq* DNA polymerase (Boehringer Mannheim) in reaction buffer consisting of 50 mM Tris-HCl (pH 8.5), 20 mM KCl, 2.5 mM $MgCl_2$, and 0.5 mg of bovine serum albumin/ml. The *Taq* enzyme was added after an initial 3-min denaturation step at 95°C (4). Thermocycling was performed in a PTC-100TM Programmable Thermal Controller (MJ Research Inc.) as follows: 1 cycle of 75°C, 1 min; 50°C, 2 min; 74°C, 2 min and 30 cycles of 95°C, 1 min; 50°C, 1 min; 74°C, 2 min, followed by an 8-min elongation at 74°C.

Subcloning PCR-amplified *fimU*. PCR-amplified *fimU* was purified from a 1% agarose gel with Sephaglas, ligated to pGEM-T according to the manufacturer's instructions (Promega Corp.), and then

transformed into *E. coli* DH5α (38).

Mapping of *fimU* on genomic restriction maps of *Salmonella* and *E. coli* strains. The *fimU* gene was mapped as previously described for the four fimbrin genes *sefA*, *agfA*, *fimA*, and *sefD* (8). The *fimU* probe, prepared by running *Eco*RI- and *Hin*dIII-digested pLU/TA 4-1 on a 1% agarose gel and purifying the fragment with Sephaglas, was labelled with [α^{32}P]dATP (Pharmacia Biotech) by nick translation. The radiolabelled *fimU* probe was hybridized to nitrocellulose blots containing *Xba*I- and *Bln*I-digested *E. coli*, *S. typhimurium*, and *S. enteritidis* genomic DNA separated by pulsed field gel electrophoresis (blots were provided by K. Sanderson and S.-L. Liu; see reference 8).

RNA extraction and Northern blot analysis. Total RNA was prepared from *S. enteritidis* 3b, 3b-122, 3b-122 pGEM-T1, and 3b-122 pLU/TA 4-1 grown statically in LB or CFA broth at 37°C for 45 h by a modification of the procedure of McCormick et al. (28) as described in Clouthier et al. (7). For *fimU* transcript analysis, the RNA was separated on a 10% polyacrylamide gel containing 8 M urea and transferred onto Hybond N^+ membranes (Amersham) with transfer buffer (0.025 M phosphate buffer [pH 6.5]) and an LKB Pharmacia semidry blotting apparatus. For *sefA* transcript analysis, the electrophoretic separation of total cellular RNA and its subsequent transfer to Hybond N^+ membranes (Amersham) were performed as described in Fourney et al. (15). The *fimU*- and *sefA*-specific probes used for Northern blot analysis were gel purified from *Eco*RI and *Hin*dIII digests of pLU/TA 4-1 and pSFA, respectively, with Sephaglas. The probes were labelled with [α^{32}P]dATP (Pharmacia Biotech) by nick translation and hybridized to the blots at 65°C for 18 h in the presence of 200 μg of herring sperm DNA (Sigma)/ml. The membranes were washed at high stringency (0.2× SSC buffer-0.1% SDS, 65°C), and the results were recorded on Kodak BioMax or X-Omat AR5 film.

SDS-PAGE and Western blot analysis. Whole-cell lysates of *S. enteritidis* 3b or clones of this strain were screened for the presence of four fimbrial types. SEF14, 18, and 21 were solubilized from whole cells with SDS-polyacrylamide gel electrophoresis (PAGE) sample buffer supplemented with 0.2 M glycine (pH 2, 100°C, 10 min), whereas SEF17 fimbriae were solubilized from whole cells with formic acid according to the method of Collinson et al. (9, 10). A portion of each culture (1 OD_{630} unit) was resuspended in 200 μl of sample buffer, and 10 μl (0.01 OD_{630} unit) was loaded per lane. Proteins in these samples were separated by SDS-PAGE, electrophoretically transferred to nitrocellulose, and screened with rabbit polyclonal anti-SEF14 (7), SEF17 (10), SEF18 (6), or SEF21 immune serum (30). Immunoreactive proteins were detected with goat anti-rabbit immunoglobulin G-alkaline phosphatase conjugates (Cedarlane) and visualized with 5-bromo-4-chloro-3-indolylphosphate and Nitro Blue Tetrazolium (Sigma).

Electron microscopy. SEF14 and SEF21 fimbriae on *S. enteritidis* 3b, 3b-122, 3b-122 pGEM-T1, and 3b-122 pLU/TA 4-1 were immunogold labelled with SEF14- or SEF21-specific rabbit polyclonal immune sera followed by incubation with protein A-15-nm-diameter gold particles (Cedarlane). Negative staining was performed as described previously (10).

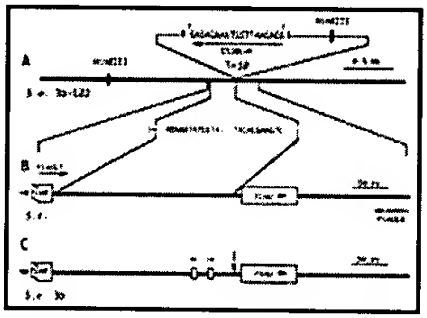
Nucleotide sequence accession number. The nucleotide sequence reported herein for *fimU* has been submitted to GenBank and has been given the accession number AF013136.

RESULTS

▲ Top
▲ Abstract
▲ Introduction
▲ Materials & Methods
▪ Results
▼ Discussion
▼ References

Fimbria production in *S. enteritidis* 3b-122. Production of SEF14, -17, -18, and -21 fimbriae by the *S. enteritidis* Tn*10* mutant 3b-122 grown under various growth conditions was assessed by Western blotting with fimbria-specific antisera, and the results were compared to those obtained with the wild-type strain *S. enteritidis* 3b. The Tn*10* mutation in 3b-122 had a pronounced effect on SEF14 and SEF21 production but little or no effect on SEF17 and SEF18 production (Table 1). As previously reported, SEF14 fimbriae were not expressed by 3b-122 grown in static CFA broth at 37°C (Table 1). Further characterization in this study, however, showed that 3b-122 lost SEF14 expression under all the growth conditions in which 3b was SEF14 positive (Table 1). In addition to the SEF14-negative phenotype, 3b-122 was also defective for type 1 fimbria (SEF21) production. The wild-type strain produced FimA under all growth conditions tested, whereas 3b-122 only produced FimA in CFA broth cultures. Thus, the result of the Tn*10* insertion was the complete loss of SEF14 expression under all growth conditions and selective loss of SEF21 expression under certain growth conditions. The altered production of SEF14 and SEF21 fimbriae in the Tn*10* insertion mutant relative to the wild-type expression patterns suggested that the transposon insertion interrupted a gene whose product was required for both SEF14 and SEF21 fimbria expression.

Identification of the Tn*10* insertion site in *S. enteritidis* 3b-122. To determine the Tn*10* insertion site, *Hin*dIII fragments of 3b-122 chromosomal DNA were subcloned into pTZ19. Clones containing Tn*10* were identified with the probe IS10L+R, which hybridized within the insertion sequence located at either end of the transposon (Fig. 1A). Of the 17 Tn*10*-positive clones identified, 3 were subjected to DNA sequence analysis. Comparison of the 3b-122 DNA sequence flanking Tn*10* to sequences listed in the NCBI databases revealed that the sequence was 99% identical to that of the region located between *fimW* and *fimU* of the *S. typhimurium* type 1 fimbrial gene cluster. Thus, on the basis of DNA sequence comparison, Tn*10* was inserted 13 bp upstream of the predicted start site of the gene, which will hereafter be referred to as *fimU* (Fig. 1A and B).

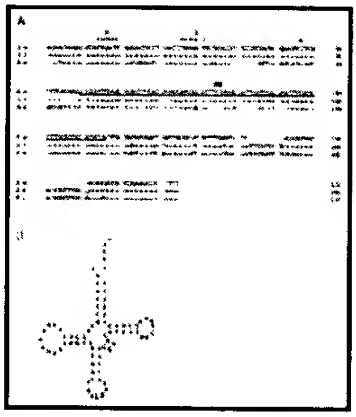


View larger version (19K):
[in this window]
[in a new window]

FIG. 1. Location of Tn*10* on the *S. enteritidis* 3b-122 chromosome and identification of the genes flanking the Tn*10* insertion in *S. enteritidis* 3b. (A) Schematic diagram of *S. enteritidis* (*S.e.*) 3b-122 chromosomal DNA (black line) showing the Tn*10* insert and the strategy used to obtain the chromosomal DNA sequence adjacent to one side of this insert. A 3-kb *Hin*dIII fragment comprising 3b-122 chromosomal DNA fused to one end of Tn*10* was identified by hybridization with the Tn*10* oligonucleotide IS10L+R. IS10L+R was also used as a sequencing primer to obtain 240 bp of DNA sequence from the subcloned *Hin*dIII fragment. (B) Schematic diagram of the *S. typhimurium* (*S.t.*) chromosome (black line) between *fimW* and *fimU* of the type I fimbrial gene cluster that was homologous to the 240 bp of *S. enteritidis* 3b-122 DNA sequence. Two 42-bp oligonucleotide primers, *fimU*LT and *fimU*LB (horizontal

arrows), were made based on the *S. typhimurium* sequence previously deposited in GenBank (L19338) by Swenson and Clegg (39). (C) Segment of the *S. enteritidis* 3b chromosome (black line) amplified by PCR with the primers *fimU*LT and *fimU*LB. This amplified DNA segment was subcloned and sequenced (Fig. 2) to identify the DNA flanking the Tn*10* insert. The Tn*10* insertion point (vertical arrow) was determined to be between the −10 region and the start of the *fimU* gene. The presence of *fimW*, *fimU*, and the −35 region on the 3b chromosome is also noted.

DNA sequence analysis of *fimU* subcloned from *S. enteritidis* 3b. By using primers *fimU*LT and *fimU*LB designed from the sequence flanking the *fimU* gene in *S. typhimurium*, a 490-bp fragment was PCR amplified from the cosmid clone pPB523-G containing 35 kb of *S. enteritidis* 3b DNA (Fig. 1B). The *fimU* PCR product was subcloned into vector pGEM-T (Fig. 1B). Nucleotide sequence analysis of three clones revealed a potential promoter, but a putative translated protein could not be detected by open reading frame analysis. Comparison of the DNA sequence downstream of the potential promoter to sequences listed in the NCBI databases showed that the sequence was identical to that of *fimU* of *S. typhimurium* (Fig. 1C and 2B) and 96% similar to that of *argU/dnaY* of *E. coli* (Fig. 2A). These genes encode arginine-specific tRNAs that recognize the rare AGA codon. The nucleotide sequence of *fimU* from *S. enteritidis* 3b contained 4 inverted repeats, which were predicted to fold the sequence into the characteristic tRNA-like cloverleaf structure (Fig. 2B) with UCU in the expected tRNA anticodon position. Together, these data suggested that *fimU* from *S. enteritidis* 3b encoded an arginine-specific tRNA.



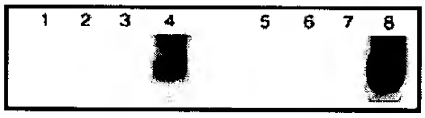
View larger version (22K):
[in this window]
[in a new window]

FIG. 2. Sequence comparison of *fimU* from *S. enteritidis* 3b (*S.e.*) with *fimU* of *S. typhimurium* (S.t.) and *argU* of *E. coli* (*E.c.*) as well as the predicted *fimU* RNA secondary structure. (A) Alignment of *S. enteritidis fimU* DNA sequence with both the *S. typhimurium fimU* (39) and *E. coli* (31) *argU* gene sequences. Symbols: •, DNA sequence identity; −, gaps introduced to maximize homology; *, bases constituting the −35 and −10 boxes; ⇓, bases constituting the anticodon; ▽, position of the Tn*10* insertion on the *S. enteritidis* 3b-122 chromosome. The DNA sequence corresponding to the proposed mature $tRNA^{Arg}$ (UCU) is underlined. (B) Diagram of the proposed secondary structure for $tRNA^{Arg}$ (UCU) from *S. enteritidis* 3b. The anticodon bases are underlined.

Analysis of the nucleotide sequence downstream of *fimU* revealed an open reading frame oriented in the opposite direction such that the 3' ends of the two genes overlapped. The predicted amino acid sequence was 88% similar to that of the prophage DLP12 integrase of *E. coli*. The sequence further downstream of *fimU* displayed 60 to 88% similarities to those encoding transposases of the IS*3* family of insertion elements.

Mapping *fimU* on the *S. enteritidis* 3b genome. Like *fimA*, *fimU* was localized to chromosomal *Xba*I and *Bln*I fragments in the 98.5- to 13.0-C's region of the chromosome in both *Salmonella* serovars. By using a series of *S. typhimurium* and *S. enteritidis* Tn*10* mutants the *fimU* gene was more precisely mapped to between *purE884*::Tn*10* at 12.6 C's and the first *Xba*I restriction site at 13.6 C's in *S. enteritidis* or 13.0 C's in *S. typhimurium*. Thus, *fimU* mapped to the same region shown previously to contain the *fimA* gene in the *fim* gene cluster.

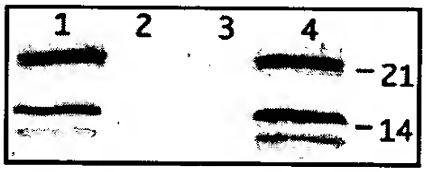
Analysis of *fimU* transcription. To determine if the *fimU* transcript was the same size as tRNA, a *fimU*-specific probe was hybridized to a blot containing total RNA isolated from *S. enteritidis* 3b, 3b-122, 3b-122 pGEM-T1, or 3b-122 pLU/TA 4-1 grown under conditions optimal for type 1 fimbria (SEF21) production in *S. enteritidis* 3b (static LB broth, 48 h, 37°C). The *fimU*-specific probe hybridized to a transcript that was present in total RNA from 3b and 3b-122 pLU/TA 4-1 (Fig. 3). The *fimU* transcript was consistently difficult to detect on Northern blots of 3b RNA even with excessive amounts of RNA loaded on the gels (25 μg [Fig. 3, lanes 1 to 4] and 50 μg [Fig. 3, lanes 5 to 8]) and extended exposure of the blots to X-ray film. Although the transcript was not found in RNA prepared from 3b-122, trace levels were evident in 3b-122 carrying the vector pGEM-T (Fig. 3), but the transcript was even more difficult to detect than its counterpart in 3b. The transcript detected with the *fimU*-specific probe comigrated with tRNA, suggesting that the product of *fimU* from *S. enteritidis* was indeed a tRNA molecule.



View larger version (29K):
[in this window]
[in a new window]

FIG. 3. Northern blot analysis of $tRNA^{Arg}$ (UCU) production in *S. enteritidis* 3b strains. A *fimU*-specific probe was hybridized to PAGE-separated total RNA from *S. enteritidis* 3b (lanes 1 and 5), 3b-122 (lanes 2 and 6), 3b-122 pGEM-T1 (lanes 3 and 7), or 3b-122 pLU/TA 4-1 (lanes 4 and 8). Lanes 1 to 4 contain 25 μg of RNA, and lanes 5 to 8 contain 50 μg of RNA.

Complementation of fimbrin expression and fimbria assembly in 3b-122. Fimbria expression was examined in *S. enteritidis* 3b, 3b-122, 3b-122 pGEM-T1, and 3b-122 LU/TA 4-1 grown under conditions that promoted production of both SEF14 and SEF21 by the wild-type strain (static CFA broth, 48 h, 37°C). Western blot analysis of whole-cell lysates using SEF14- or SEF21-specific antisera showed that complementation of the insertion mutation in 3b-122 with pLU/TA 4-1 restored SEF14 and SEF21 fimbria expression (Fig. 4). Thus *fimU* affected the production of two fimbrins encoded by genes located on two different gene clusters.



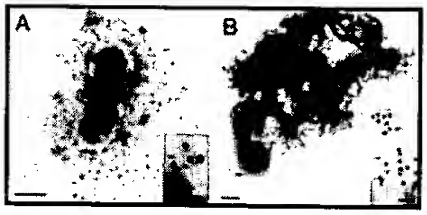
View larger version (49K):
[in this window]
[in a new window]

FIG. 4. Complementation analysis of *S. enteritidis* 3b-122 Tn*10* mutant with the *fimU*-containing recombinant plasmid pLU/TA 4-1. Whole-cell extracts were analyzed by Western blotting to determine the presence of SefA (21 kDa) and FimA (14 kDa) fimbrin proteins in *S. enteritidis* 3b (lane 1), 3b-122 (lane 2), 3b-122 pGEM-T1 (lane 3), and 3b-122 pLU/TA 4-1 (lane 4). Numbers at right indicate positions of SefA (21) and FimA (14).

Assembly of SEF14 and SEF21 fimbriae on the cell surface of *S. enteritidis* 3b, 3b-122, 3b-122 pGEM-T1, or 3b-122 pLU/TA 4-1 was determined by immunogold labelling and electron microscopy performed on cells grown in static CFA broth for 48 h at 37°C. The wild-type strain, *S. enteritidis* 3b, expressed both SefA and FimA and assembled the respective subunits into SEF14 and SEF21 fimbriae (Table 2). Similar analyses of 3b-122 and 3b-122 pGEM-T1 showed that SEF14 was not produced and that SEF21 fimbriae on the cell surfaces of these two strains were rarely detected (Table 2), a result consistent with the Western blot data (Fig. 4). In contrast, SEF14 and SEF21 fimbriae were evident on the surface of 3b-122 pLU/TA 4-1 (Fig. 5) at levels equal to or greater than that on 3b. Thus, expression of SefA and FimA fimbrins and assembly of their respective fimbriae were restored by complementation of the Tn*10 fimU* mutation in 3b-122 with a wild-type copy of the *fimU* gene on pLU/TA 4-1. Cells producing SEF17 were readily seen without immunolabelling on all the grids prepared for electron microscopy (Table 2).

View this table: [in this window] [in a new window]

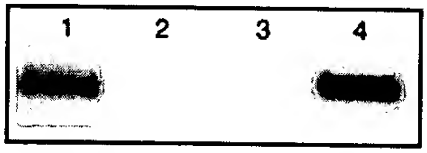
TABLE 2. Detection of assembled SEF14 and SEF21 fimbriae in various *S. enteritidis* strains by immunoelectron microscopy with specific antifimbrial sera



View larger version (90K): [in this window] [in a new window]

FIG. 5. Analysis of SEF14 and SEF21 fimbria assembly in *S. enteritidis* 3b-122 pLU/TA 4-1 by immunogold electron microscopy. *S. enteritidis* 3b-122 pLU/TA 4-1 was labeled with protein A-gold and negatively stained following incubation with immune serum generated to SEF14 (A) or SEF21 (B). Arrows indicate individual immunogold-labeled SEF14 and SEF21 fimbriae in panel A and B insets, respectively. The average diameter of the gold particles was 15 nm. Bar, 0.5 μm (electron micrograph) or 50 nm (inset).

Analysis of *sefA* transcription. The effect of $tRNA^{Arg}$ (UCU) on *sefA* transcript production was analyzed by hybridizing a *sefA*-specific probe to a blot containing total cellular RNA isolated from *S. enteritidis* 3b, 3b-122, 3b-122 pGEM-T1, or 3b-122 LU/TA 4-1 grown under conditions optimal for SEF14 production in 3b (static CFA broth, 48 h, 37°C). The *sefA*-specific probe hybridized to a 660-base transcript that was present in RNA from 3b and 3b-122 pLU/TA 4-1 but absent from RNA from 3b-122 and 3b-122 pGEM-T1 (Fig. 6). The strains expressing the *sefA* transcript corresponded to those carrying a functional *fimU* gene, suggesting that *fimU* was required for expression of *sefA*.



View larger version (36K): [in this window] [in a new window]

FIG. 6. Northern blot analysis of *sefA* transcription in *S. enteritidis* 3b strains. A *sefA*-specific probe was hybridized to 10 μg of total RNA from *S. enteritidis* 3b (lane 1), 3b-122 (lane 2), 3b-122 pGEM-T1 (lane 3), or 3b-122 LU/TA 4-1 (lane 4).

DISCUSSION

▲ Top
▲ Abstract
▲ Introduction
▲ Materials & Methods
▲ Results
• Discussion
▼ References

fimU, located in the *fim* (*sef21*) operon of *S. enteritidis* 3b, encodes an arginine-specific tRNA that is required for expression of not only SEF21 fimbriae (type 1) but also SEF14 fimbriae. The product of *fimU* in 3b is a tRNA, since Northern blot analysis of RNA from 3b and 3b-122 pLU/TA 4-1 demonstrates that the *fimU* transcript comigrates on polyacrylamide gels with tRNA. Furthermore, the 77-nucleotide sequence of the *fimU* gene of 3b is identical to that of *fimU* of *S. typhimurium* (39) and shares extensive homology with that of *argU* encoding $tRNA^{Arg}$ (UCU) from *E. coli* (18). The *fimU*-encoded transcript from 3b can be folded into a typical tRNA cloverleaf structure containing the 3'-terminal sequence CCA as well as the invariant or semivariant nucleotides common to tRNA molecules (19, 34). Finally, the DNA sequence 5' to the *fimU* gene contains features common to the promoters of tRNA operons including the consensus *E. coli* −10 and −35 promoter elements (16, 21) and a G+C-rich discriminator sequence (16, 42, 43). The regulatory mechanisms controlling *fimU* expression are unknown. Recently, however, the integration and excision of plasmids, phage, and pathogenicity islands into and out of the chromosomes at tRNA loci have been shown to affect tRNA gene expression (20, 33). As shown with the *E. coli* tRNA gene *argU* (25), the open reading frame adjacent to and overlapping *fimU* is a homolog of the integrase gene (*int*) from the defective lambdoid prophage DLP12. Integration of prophage DLP12 at this site prevents cotranscription of the *int* gene with *fimU*, which may contribute to the regulation of *fimU* expression in *S. enteritidis* 3b.

The influence of $tRNA^{Arg}$ (UCU) encoded by *fimU* on SEF14 and SEF21 fimbria production is evident in the Tn*10* insertion mutant *S. enteritidis* 3b-122. The transposon, inserted between the predicted promoter and the 5' end of the mature *fimU* transcript, disrupts transcription of *fimU* and thus $tRNA^{Arg}$ (UCU) production. This mutation results in the loss of SefA production and selective loss of FimA production, i.e., the subunits of SEF14 and SEF21 (type 1) fimbriae, respectively. Thus, in *S. enteritidis* 3b, $tRNA^{Arg}$ (UCU) is required for SEF14 production and enhances type 1 fimbria (SEF21) production. In *E. coli*, cross-talk has also been reported to occur between adhesin gene clusters (29), and tRNA molecules have been shown to play a key role in global regulatory cascades (20). However, this is the first study to show that a tRNA-specific locus found on one fimbrial operon influences the production of two fimbrins whose operons are separated by 15 C's on the chromosome.

$tRNA^{Arg}$ (UCU) encoded by *fimU* is required for transcription of *sefA*, the gene encoding the subunit of SEF14 fimbriae in *S. enteritidis* 3b. The regulatory mechanism is unknown, but a direct correlation between the abundance of tRNAs and the occurrence of the respective codons in protein genes (22, 23) has been suggested to control the translation of genes containing rare codons (3, 37). Since AGA, the codon recognized by the $tRNA^{Arg}$ (UCU) species encoded by *fimU*, is one of the least-used codons for arginine, then perhaps the limited availability of charged tRNAs for this minor codon controls the level of translation of the *sefA* transcript or of a transcript whose protein product is involved in the regulation

of *sefA* transcription. *sefA* contains neither of the rare arginine codons AGA or AGG recognized by $tRNA^{Arg}$ (UCU), indicating that *fimU* expression would not have a direct effect on the translation of *sefA* mRNA. However, *sefE*, the gene encoding the putative AraC-like transcriptional activator of the *sef14* gene cluster, contains 13 arginine codons, including 9 AGA codons and 1 AGG codon (5). Perhaps the $tRNA^{Arg}$ (UCU) encoded by *fimU* regulates translation of *sefE*, which would in turn affect transcription of *sefA* and the downstream genes.

With the exception of the gene *fimA* encoding the subunit of SEF21 fimbriae (12), the remainder of the *sef21* gene cluster has not been characterized in *S. enteritidis* 3b. Thus, the mechanism for regulation of type 1 fimbria synthesis by *fimU* remains to be determined. However, type 1 fimbria (SEF21) production is optimal when *S. enteritidis* 3b is grown at 37°C in a static broth culture but suboptimal when the cells are grown at lower temperatures (21 to 30°C) in shaking broth culture or on solid medium (12). Similarly, expression of SEF14 fimbriae by *S. enteritidis* 3b is environmentally controlled by temperature, medium composition, and aeration, and is optimal at 37°C in static, aerobic CFA broth (5). Thus, the coregulation of SefA and FimA fimbrin production by *fimU*-encoded $tRNA^{Arg}$ (UCU) results in the corresponding fimbriae being expressed under similar environmental conditions, which may give the bacteria a competitive advantage for survival.

ACKNOWLEDGMENTS

This project was supported by operating grants to W.W.K. provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Bacterial Diseases Network. S.C.C. was the recipient of an Industrial Postdoctoral Fellowship from NSERC, and A.P.W. was the recipient of a Graduate Research Engineering and Technology Award scholarship from the Science Council of British Columbia and a Graduate Research scholarship from NSERC.

We thank S. Lee for helpful discussions on tRNA.

FOOTNOTES

* Corresponding author. Mailing address: Department of Biochemistry and Microbiology, University of Victoria, Victoria, B.C. V8W 3P6, Canada. Phone: (250) 721 7078. Fax: (250) 721 8882. E-mail: wkay@uvic.ca.

REFERENCES

▲ Top
▲ Abstract
▲ Introduction
▲ Materials & Methods
▲ Results
▲ Discussion
▪ References

1. **Alm, R. A., P. Guerry, and T. J. Trust.** 1993. Distribution and polymorphism of the flagellin

genes among isolates of *Campylobacter coli* and *Campylobacter jejuni*. J. Bacteriol. **175:**3051-3057 [Abstract].

2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403-410[Medline].
3. **Chen, G.-F., and M. Inouye.** 1994. Role of the AGA/AGG codons, the rarest codons in global gene expression in *Escherichia coli*. Genes Dev. **8:**2641-2652[Abstract].
4. **Chou, Q., M. Russell, D. E. Birch, J. Raymond, and W. Bloch.** 1992. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplification. Nucleic Acids Res. **20:**1717-1723[Abstract].
5. **Clouthier, S. C.** 1995. . Ph.D. thesis. University of Victoria, Victoria, Canada.
6. **Clouthier, S. C., S. K. Collinson, and W. W. Kay.** 1994. Unique fimbriae-like structures encoded by *sefD* of the SEF14 fimbrial gene cluster of *Salmonella enteritidis*. Mol. Microbiol. **12:**893-903 [Medline].
7. **Clouthier, S. C., K.-H. Müller, J. L. Doran, S. K. Collinson, and W. W. Kay.** 1993. Characterization of three fimbrial genes, *sefABC*, of *Salmonella enteritidis*. J. Bacteriol. **175:**2523-2533[Abstract].
8. **Collinson, S. K., S.-L. Liu, S. C. Clouthier, P. A. Banser, J. L. Doran, K. E. Sanderson, and W. W. Kay.** 1996. The location of four fimbrin-encoding genes, *agfA*, *fimA*, *sefA*, and *sefD*, on the *Salmonella enteritidis* and/or *S. typhimurium Xba*I-*Bln*I genomic restriction maps. Gene **169:**75-80 [Medline].
9. **Collinson, S. K., P. C. Doig, J. L. Doran, S. C. Clouthier, T. J. Trust, and W. W. Kay.** 1993. Thin aggregative fimbriae mediate binding of *Salmonella enteritidis* to fibronectin. J. Bacteriol. **175:**12-18[Abstract].
10. **Collinson, S. K., L. Emödy, K.-H. Müller, T. J. Trust, and W. W. Kay.** 1991. Purification and characterization of thin, aggregative fimbriae from *Salmonella enteritidis*. J. Bacteriol. **173:**4773-4781[Medline].
11. **Doran, J. L., S. K. Collinson, S. C. Clouthier, T. A. Cebula, W. H. Koch, J. Burian, P. A. Banser, E. C. D. Todd, and W. W. Kay.** 1996. Diagnostic potential of *sefA* DNA probes to *Salmonella enteritidis* and certain other O-serogroup D1 *Salmonella* serovars. Mol. Cell. Probes **10:**233-246[Medline].
12. **Doran, J. L., S. K. Collinson, C. M. Kay, P. A. Banser, J. Burian, C. K. Munro, S. H. Lee, J. M. Somers, E. C. D. Todd, and W. W. Kay.** 1994. *fimA* and *tct* based DNA diagnostics for *Salmonella*. Mol. Cell. Probes **8:**291-310[Medline].
13. **Evans, D. G., D. J. Evans, Jr., and W. Tjoa.** 1977. Hemagglutination of human group A erythrocytes by enterotoxigenic *Escherichia coli* isolated from adults with diarrhea: correlation with colonization factor. Infect. Immun. **18:**330-337[Medline].
14. **Feutrier, J., W. W. Kay, and T. J. Trust.** 1986. Purification and characterization of fimbriae from *Salmonella enteritidis*. J. Bacteriol. **168:**221-227[Medline].
15. **Fourney, R. M., J. Miyakoshi, R. S. Kay III, and M. C. Paterson.** 1992. Northern blotting: efficient RNA staining and transfer. Focus **10:**5-7.
16. **Fournier, M. J., and H. Ozeki.** 1985. Structure and organization of the transfer ribonucleic acid genes of *Escherichia coli* K-12. Microbiol. Rev. **49:**379-397.
17. **Gally, D. L., J. Leathart, and I. C. Blomfield.** 1996. Interaction of FimB and FimE with the *fim* switch that controls the phase variation of type 1 fimbriae in *Escherichia coli* K-12. Mol. Microbiol. **21:**725-738[Medline].
18. **Garcia, G. M., P. K. Mar, D. A. Mullin, J. R. Walker, and N. E. Prather.** 1986. The *E. coli dnaY* gene encodes an arginine transfer RNA. Cell **45:**453-459[Medline].

19. **Gauss, D. H., and M. Sprinzl.** 1983. Compilation of tRNA sequences. Nucleic Acids Res. **11:**1-53 [Medline].

20. **Hacker, J., G. Blum-Oehler, I. Mühldorfer, and H. Tschäpe.** 1997. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. Mol. Microbiol. **23:**1089-1097[Medline].

21. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11:**2237-2255[Abstract].

22. **Ikemura, T.** 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for synonymous codon choice that is optimal for the *E. coli* translational system. J. Mol. Biol. **151:**389-409[Medline].

23. **Ikemura, T.** 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. J. Mol. Biol. **146:**1-21[Medline].

24. **Klemm, P.** 1986. Two regulatory *fim* genes, *fimB* and *fimE*, control the phase variation of type 1 fimbriae in *Escherichia coli*. EMBO J. **5:**1389-1393[Abstract].

25. **Lindsey, D. F., D. A. Mullin, and J. R. Walker.** 1989. Characterization of the cryptic lambdoid prophage DLP12 of *Escherichia coli* and overlap of the DLP12 integrase gene with the tRNA gene *argU*. J. Bacteriol. **171:**6197-6205[Medline].

26. **Marck, C.** 1988. "DNA Strider": a "C" program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucleic Acids Res. **16:**1829-1836[Abstract].

27. **McClain, M. S., I. C. Blomfield, and B. I. Eisenstein.** 1991. Roles of *fimB* and *fimE* in site-specific DNA inversion associated with phase variation of type 1 fimbriae in *Escherichia coli*. J. Bacteriol. **173:**5308-5314[Medline].

28. **McCormick, J. R., J. M. Zengel, and L. Lindahl.** 1991. Intermediates in the degradation of mRNA from the lactose operon of *Escherichia coli*. Nucleic Acids Res. **19:**2767-2776[Abstract].

29. **Morschhauser, J., V. Vetter, L. Emody, and J. Hacker.** 1994. Adhesin regulatory genes within large, unstable DNA regions of pathogenic *Escherichia coli*: cross-talk between different adhesin gene clusters. Mol. Microbiol. **11:**555-566[Medline].

30. **Müller, K.-H., S. K. Collinson, T. J. Trust, and W. W. Kay.** 1991. Type 1 fimbriae of *Salmonella enteritidis*. J. Bacteriol. **173:**4765-4772[Medline].

31. **Mullin, D. A., G. M. Garcia, and J. R. Walker.** 1984. An *E. coli* DNA fragment 118 base pairs in length provides *dnaY*$^{+}$ complementing activity. Cell **37:**669-674[Medline].

32. **Newman, J. V., R. L. Burghoff, L. Pallesen, K. A. Krogfelt, C. S. Kristensen, D. C. Laux, and P. S. Cohen.** 1994. Stimulation of *Escherichia coli* F-18Col^{-} type-1 fimbriae synthesis by *leuX*. FEMS Microbiol. Lett. **122:**281-287[Medline].

33. **Reiter, W.-D., P. Palm, and S. Yeats.** 1989. Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. Nucleic Acids Res. **17:**1907-1914[Abstract].

34. **Rich, A., and U. L. RajBhandary.** 1976. Transfer RNA: molecular structure, sequence, and properties. Annu. Rev. Biochem. **45:**805-860[Medline].

35. **Ritter, A., D. L. Gally, P. B. Olsen, U. Dobrindt, A. Friedrich, P. Klemm, and J. Hacker.** 1997. The Pai-associated *leuX* specific tRNA^{Leu} affects type 1 fimbriation in pathogenic *Escherichia coli* by control of FimB recombinase expression. Mol. Microbiol. **25:**871-882[Medline].

36. **Rosner, J. L.** 1972. Formation, induction and curing of bacteriophage P1 lysogens. Virology **48:**679[Medline].

37. **Saier, M. H., Jr.** 1995. Differential codon usage: a safeguard against inappropriate expression of specialized genes? FEBS Lett. **362:**1-4[Medline].

38. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. . Molecular cloning: a laboratory manual,

2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

39. **Swenson, D. L., and S. Clegg.** 1994. *Salmonella* fimbriae, p. 105-113. *In* P. Klemm (ed.), Fimbriae: adhesion, biogenesis, genetics and vaccines. CRC Press, London, United Kingdom.
40. **Swenson, D. L., K.-J. Kim, E. W. Six, and S. Clegg.** 1994. The gene *fimU* affects expression of *Salmonella typhimurium* type 1 fimbriae and is related to the *Escherichia coli* tRNA gene *argU*. Mol. Gen. Genet. **244:**216-218[Medline].
41. **Swenson, D. L., and S. Clegg.** 1992. Identification of ancillary *fim* genes affecting *fimA* expression in *Salmonella typhimurium*. J. Bacteriol. **174:**7697-7704[Abstract].
42. **Travers, A. A.** 1984. Conserved features of coordinately regulated *E. coli* promoters. Nucleic Acids Res. **12:**2605-2618[Abstract].
43. **Travers, A. A.** 1980. Promoter sequence for stringent control of bacterial ribonucleic acid synthesis. J. Bacteriol. **141:**973-976[Medline].

J Bacteriol, February 1998, p. 840-845, Vol. 180, No. 4
0021-9193/98/$04.00+0

This article has been cited by other articles:

- Edwards, R. A., Matlock, B. C., Heffernan, B. J., Maloy, S. R. (2001). Genomic analysis and growth-phase-dependent regulation of the SEF14 fimbriae of Salmonella enterica serovar Enteritidis. *Microbiology* 147: 2705-2715 [Abstract] [Full Text]
- Tinker, J. K., Hancox, L. S., Clegg, S. (2001). FimW Is a Negative Regulator Affecting Type 1 Fimbrial Expression in Salmonella enterica Serovar Typhimurium. *J. Bacteriol.* 183: 435-442 [Abstract] [Full Text]
- Tinker, J. K., Clegg, S. (2000). Characterization of FimY as a Coactivator of Type 1 Fimbrial Expression in Salmonella enterica Serovar Typhimurium. *Infect. Immun.* 68: 3305-3313 [Abstract] [Full Text]
- Julio, S. M., Heithoff, D. M., Mahan, M. J. (2000). ssrA (tmRNA) Plays a Role in Salmonella enterica Serovar Typhimurium Pathogenesis. *J. Bacteriol.* 182: 1558-1563 [Abstract] [Full Text]